A Publication of the American Psychological Association

The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

Conditioning and Personality: A Study of Normal and Neurotic Subjects.... 143
Cyril M. Franks

Self-Appraisal, Personal Desirability, and Perceived Social Desirability of Personality Traits..... 151
Ephraim Rosen

Discrimination and Reliability in Q-Sort Personality Descriptions..... 159
Norman H. Lisson and Thomas F. Nichols

The Role of Small Groups in Mediating the Effects of Propaganda..... 166
May Brodbeck

Self Ratings of Fear in a Fear-Invoking Situation..... 171
Richard D. Walk

A Factorial Study of Adjustment to Stress..... 179
Wayne H. Holtzman and M. E. Bitterman

Motivation and Performance in Schizophrenia..... 186
Bertram D. Cohen

Time and Space Orientation in Schizophrenics..... 191
C. O. de la Garza and Philip Worchel

Intra-Individual Variability in Sentence Completion Responses..... 195
Jerry Osterweil and Donald W. Fiske

Continued on inside front cover

Volume 52 MARCH 1956 Number 2

CONTENTS—*Continued*

Intertrial Effects of Immediate Self-Committal in a Continuous Social Influence Situation 200
Seymour Fisher, Irvin Rubinstein, and Robert W. Freeman
Abstract Thinking and Social Behavior in Schizophrenia 208
John H. Flavell
Self-Consistency as a Factor Affecting Immediate Recall 212
Desmond S. Cartwright
Social Status and Success Strivings 219
Elizabeth Dowdan
An Experimental and Theoretical Analysis of Perceptual Defense 224
Charles W. Eriksen and C. Thayer Browne
Object Constancy in Schizophrenia: The Enhancement of Symbolic Objects and Conceptual Stability 231
Harold L. Raush
Some Factors Affecting an Individual's Estimate of his Probable Success in a Group Situation 235
Harold B. Gerard
Future Time Perspective in Schizophrenia 240
Melvin Wallace
Mother-Concepts and Judgments of Young Women's Faces 246
Paul F. Secord and Sidney M. Jourard
A Comparison of Reproduction Errors Made by Brain-Damaged and Control Patients on a Memory-for-Designs Test 251
H. J. Wahler
Attitude Change Through Modification of Attitude Structure 256
Earl R. Carlson
The Reliability of Psychiatric Diagnosis: A New Look 262
Hermann O. Schmidt and Charles P. Fonda

CASE REPORTS

Motor Manifestation of Conflict in Interview: A Case Study 268
Robert B. Malmo, A. Arthur Smith, and Werner A. Kohlmeyer
A Case Study of a Multiple Personality 272
V. K. Alexander

CRITIQUE AND NOTES

The Use of Personal Pronouns as Categories for Studying Small Group Interaction 277
Dorothy Clifton Conrad and Richard Conrad
An Example of Assimilative Projection 279
Elizabeth Minin
Responses to a Personality Questionnaire by Volunteers and Nonvolunteers to a Kinsey Interview 280
Aaron W. Siegman
The Use of the MMPI for Cross-Cultural Personality Study: A Preliminary Report on the German Translation 281
Norman D. Sundberg
Grouping Error and Interpersonal Affect: A Correction 283
Everett W. Bovard, Jr.

Volume 52 MARCH 1956 Number 2

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, Social Science Research Council

Associate Editor: E. J. SHOBEN, JR., Teachers College, Columbia University

Managing Editor: LORRAINE BOUTHILET

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Center for Advanced Study in the Behavioral Sciences

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
The RAND Corporation

RICHARD CHRISTIE
New York University

MORTON DEUTSCH
New York University

I. E. FARBER
State University of Iowa

LEON FESTINGER
Center for Advanced Study in the Behavioral Sciences

ROBERT R. HOLT
New York University

HERBERT H. HYMAN
Columbia University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
University of Minnesota

GARDNER LINDZEY
Center for Advanced Study in the Behavioral Sciences

DAVID C. MCCLELLAND
Wesleyan University

QUINN MCNEMAR
Stanford University

LEO POSTMAN
University of California

NEVITT SANFORD
Vassar College

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.
AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6. D. C.

Entered as second-class matter February 15, 1954 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 34.40 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1956 by the American Psychological Association, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis; those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area, this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 784 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

DR. M. BREWSTER SMITH
Social Science Research Council
230 Park Avenue
New York 17, N. Y.

Fifty free offprints are supplied to contributors. Authors of prior publication articles receive no gratis offprints.

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 15th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Printed and mailed by:

WAVERLY PRESS, INC.
Baltimore 2, Maryland

CONDITIONING AND PERSONALITY: A STUDY OF NORMAL AND NEUROTIC SUBJECTS[1]

CYRIL M. FRANKS

University of London, Institute of Psychiatry, Maudsley Hospital

PAVLOV's theory of cortical functioning (24, 25, 26) emphasizes two basic cortical processes, *excitation* and *inhibition*. The term *inhibition* is unfortunately used in at least three different senses. There is the general psychiatric usage of the word, as applied to the introverted, withdrawn personality; there is the neurophysiological use of the word; there is the Pavlovian usage. It must be stressed that cortical inhibition in the Pavlovian sense (which is how it is used in the present paper) should be associated with the *absence* of behavioral inhibition in the psychiatric sense. Pavlov (24, 25, 26) has produced ample evidence to suggest that both *excitation* and *inhibition* are positive and molar cortical processes. However, any further assumptions about the physiological nature of these two processes would be largely speculative; it must be sufficient for the present to regard them both as hypothetical constructs (23) and not merely as intervening variables.

During his experiments, Pavlov observed large individual differences in the behavior of his dogs. Certain dogs (which he called the excitatory kind) developed stable positive conditioned responses with ease and retained these responses for a long time during extinction. Other dogs (which he called the inhibitory kind) developed positive conditioned responses very poorly, which, once formed, were easily disrupted and soon extinguished. After a discussion of experimentally produced "neuroses" in animals and in man, Pavlov wrote as follows:

> It has been seen that the above-mentioned method may lead to different forms of disturbance, depending on the type of nervous system of the animal. In dogs with the more resistant nervous system it leads to a predominance of excitation; in dogs with the less resistant nervous system, to a predominance of inhibition. So far as can be judged on the basis of casual observation I believe that these two variations in the pathological disturbance of the cortical activity in animals are comparable to the two forms of neurosis in man—in the pre-Freudian terminology *neurasthenia* and *hysteria*—the first with exaggeration of the excitatory and weakness of the inhibitory process, the second with a predominance of the inhibitory and weakness of the excitatory process (24, p. 397f.).

Pavlov never followed up this observation, and it has been neglected ever since. From this observation, however, it is possible to predict that neurotics of the neurasthenic type should form conditioned reflexes rapidly and that these reflexes should be slow to extinguish, whereas neurotics of the hysteric type should form conditioned reflexes slowly and that these reflexes should extinguish readily. Since the term *neurasthenic* is ill defined and rarely used in contemporary psychology, it would seem better to examine both *neurasthenia* and *hysteria* more carefully in an attempt to discover how they differ rather than to test the above prediction immediately. Both neurasthenia and hysteria are generally regarded as forms of neuroses, so it is along a dimension other than neuroticism that differences must be sought. Pavlov himself has implicitly suggested the dimension of excitation—inhibition, the neurasthenic being at the excitation end of the continuum and the hysteric at the inhibition end.

In seeking another dimension, Eysenck pointed out (1, p. 52f.) that Jung (21, p. 421) suggested that the characteristic neurosis of the extravert is hysteria, whereas that of the introvert is psychasthenia. Jung also emphasized the independence of introversion and neuroticism, and this independence has been experimentally confirmed by Eysenck and his co-workers (1, 2, 3). They have demonstrated the existence of two orthogonal factors, introversion-extraversion and neuroticism, and have shown that anxiety neurotics, obsessive-compulsives, and depressives typically have high scores on tests of neuroticism *and* high scores on tests of introversion (with correspondingly low scores on tests of extraversion). To these introverted neurotics Eysenck has given the name of dysthymics. He has also shown that hysterics and psychopaths have high scores on tests of neuroticism and high scores on tests of extraversion (with correspondingly low scores on tests of introversion).

[1] This research has been aided by a grant from the Central Research Fund of the University of London.

The neurasthenic type of neurotic would seem to be subsumed under the heading of *dysthymia* and the hysteric under the heading of *hysteria* (used here in Eysenck's sense to include both hysteria and psychopathy). Furthermore, an examination of patients usually included in these categories suggests that dysthymia is related to Pavlov's concept of excitation and hysteria to inhibition. Thus, the dysthymic usually presents overt physical and mental characteristics of anxiety, obsessions, compulsions, or ruminations. He may be irritable, introspective, ill at ease with people, very much aware of himself and of others. He is often overconscientious and highly sensitive to his environment. He may be agitated, hyperactive, and tense, or he may be reflective in his thought, overcautious and hesitant. All these characteristics are consistent with a presumed state of cortical excitation.

The hysteric group presents very different characteristics. They may develop "escape mechanisms" such as fugues, amnesia, or gross conversion symptoms. They tend to be insensitive, irresponsible, unreliable, and little concerned about other people. Most of these characteristics seem to involve some form of dissociation and may hence be reasonably conceived as associated with a state of cortical inhibition.

Pavlov's original observation may, therefore, be slightly revised in the form of the following hypothesis:

Neurotics of the dysthymic type form conditioned reflexes rapidly, and these reflexes are difficult to extinguish; neurotics of the hysteric type form conditioned reflexes slowly, and these reflexes are easy to extinguish.

PREVIOUS STUDIES

There have been many experimental studies on the conditionability of neurotics (e.g., 5, 31). Most workers have found that "neurotics" as a group condition better than normals. They have usually made no attempt, however, to classify the neurotics into dysthymics and hysterics. Since more patients are of the dysthymic variety than the hysteric, these findings are probably more accurately interpreted as indicating that dysthymic neurotics condition better than normals. More recently, Taylor and Spence (35) compared the conditionability of anxious neurotics with "other neurotics" and failed to find any statistically significant differences. It is significant, however, that the diagnoses represented in the general category of "other neurotics" included reactive depression, alcoholism, psychopathy, character disorders, obsessive compulsions, compulsions, etc. In other words, this classification of "other neurotics" included large numbers of both the dysthymic and hysteric types of neurotics. Thus, the relative conditionability of dysthymic as compared with hysteric neurotics remains an open question.

Many studies, however, have been made of the conditionability of *one kind* of dysthymia, the anxiety states. Using the PGR, for example, it has been shown that anxious subjects condition better than normals (28, 36, 37), and similar results have been obtained with the eyeblink reflex (18, 29, 30, 32).

Thus, since dysthymics and hysterics appear to differ along the dimension of introversion-extraversion, and since it seems reasonable to associate dysthymia with excitation and hysteria with inhibition, it is probable that excitation is associated with introversion and inhibition with extraversion. If this argument is sound, then conditionability is more a function of introversion-extraversion than of neuroticism. None of the studies cited bears on this hypothesis since they were concerned with the conditionability of only one group, the anxiety states. Such *S*s would form conditioned responses more readily than nonanxious *S*s either because they were more neurotic or because they were more introverted. A crucial experiment would have to examine the conditionability of both dysthymics and hysterics, and also of normals. If the ease of conditionability of the dysthymics is related to their neuroticism, then hysterics and dysthymics should condition equally well, and the normals should condition less well than either neurotic group. If, however, the ease of conditionability of the dysthymics is related to their degree of introversion (i.e., to their degree of excitation), then it follows that the hysteric group would condition the least, the dysthymic group the most, and the normals in between. Furthermore, if conditionability is related to introversion-extraversion, then within the normal group there should be no correlation between conditionability and a test of neuroticism, but a significant negative corre-

lation between conditionability and a test of extraversion.

THE PRESENT EXPERIMENT

Twenty dysthymic patients, 20 hysteric patients, and 20 normals were tested in a specially constructed soundproof conditioning laboratory (7). The unconditioned stimulus (US) was a puff of air, the conditioned stimulus (CS) a tone administered through a pair of headphones, and the unconditioned (UR) and conditioned (CR) responses were both eyeblinks and PGR changes. Partial conditioning was used, each *S* being given 30 reinforcements, interspersed with 18 test trials, and 10 extinction trials. All conditioning was carried out in one session, taking approximately half an hour per *S*. The 20 normal *S*s were retested after an interval of 14–21 days. All *S*s were given personality questionnaires defining neuroticism and introversion-extraversion, together with Taylor's Manifest Anxiety Scale (33) and the Maudsley Medical Questionnaire (MMQ).

Subjects

All *S*s were aged between 17 and 47 years. Although the normal group was significantly younger than the two neurotic groups, the neurotic groups did not differ significantly in age. The groups were not matched for sex since there is some evidence (22) to suggest that there is no relationship between sex and eyeblink conditionability, and there are no reasons for suspecting that sex differences exist with respect to PGR conditioning. According to Hilgard and Marquis (19, p. 303), there is little correlation between intelligence and conditionability; consequently, no attempt was made to match the three groups on this variable other than to exclude any *S*s of low intelligence, the criterion being a raw score of less than 29[2] on the Raven Matrices nonverbal test of intelligence (27). All three groups happen to be significantly different from each other in intelligence. The data relating to age, sex, intelligence, and extraversion are presented in Table 1.

The 20 normal subjects. The normal *S*s consisted of Occupational Therapy students, student male nurses, and Occupation Centre student teachers. All were naive with respect to psychology, and all were volunteers. They were requested to avoid discussing the experiment with friends. Twenty-three *S*s were originally tested, of whom one was rejected because she blinked excessively; one was rejected because he did not know sufficient English to answer the personality questionnaires, and one was rejected because he had recently been discharged from a mental hospital.

The 40 neurotic subjects (20 dysthymics, 20 hysterics). All the neurotic *S*s were residential or outpatients at the Maudsley Hospital. No patient was used who had any evidence or history of psychotic features, brain injury, or epilepsy, or who had received any form of psychosurgery. No *S* was used who had any physical treatment such as ECT or insulin during the previous eight weeks or who had been recently receiving such medication as sodium amytal. (All *S*s were denied any drugs for the preceding 24 hours.)

TABLE 1
DATA ON AGE, SEX, INTELLIGENCE, AND EXTRAVERSION SCORES
(N = 20 in each group)

Group	Age in Years		Sex (No. of Subjects)		Matrices Raw Score		*R* Score	
	Mean	*SD*	M	F	Mean	*SD*	Mean	*SD*
Hysterics	28.8	7.1	7	13	44.6	9.0	49.6	11.2
Normals	24.3	4.9	6	14	57.0	3.0	40.5	11.6
Dysthymics	31.3	8.2	15	5	50.2	6.7	20.6	6.3

The *S*s were included in the hysteric group if the psychiatrist concerned could diagnose them as having one or more of the following characteristics: hysterical personality, conversion symptoms, hysteria, psychopathic personality, psychopathic features. Furthermore, the psychiatrist was requested to select those *S*s with a minimum of obsessive-compulsive features, manifest anxiety, or depression.

The *S*s were included in the dysthymic group if they could be diagnosed as having one or more of the following characteristics: manifest anxiety, reactive depression, obsessive-compulsive features. They also had to show a minimum of hysteric and psychopathic characteristics.

One final criterion was used in selecting the two neurotic groups. All hysteric patients with a score of less than 33 on Guilford's *R* Scale (13) were rejected, and all dysthymic patients having an *R* score of above 32 were also rejected. This criterion was included since it has been shown (3, 17) that Guilford's *R* Scale is one of the best available measures of extraversion. Thus, the two neurotic groups were deliberately chosen to minimize overlap along the dimension of introversion-extraversion. By this means, any differences in conditionability between introverted and extraverted neurotics should become more readily apparent with a relatively small number of cases. The means and *SD*'s for the three neurotic and normal groups are presented in Table 1. Although all three groups are significantly different from each other with respect to their *R* scores, the normal group tends to resemble the hysterics more than they do the dysthymics in this respect.

Twenty-eight hysterics patients were originally tested, of whom eight had to be rejected because seven had *R* scores below 33, and because of a major error by the experimenter during conditioning with one. The final group of twenty hysterics were diagnostically constituted as follows: eight, hysteria; two, hysterical features in an immature personality; two, conversion hysteria; one, anorexia nervosa hysteria; one, hysteria plus mixed neurotic reactions; one, hysterical palsy; three, sexual psychopathy; one, chronic alcoholism in a psychopathic personality; one, delinquency in a psychopathic personality.

[2] In terms of the Wechsler-Bellevue scale, this would mean very approximately that *S*s with an IQ below 90 would be excluded. However, concern here was not with an *S*'s IQ, but with his absolute amount of intelligence, his raw score on the test concerned irrespective of age, sex, etc.

Twenty-seven dysthymic patients were originally tested of whom seven had to be rejected for the following reasons: four had *R* scores above 32; two had changed diagnoses; one was initially sensitive to the sound, giving both a blink and a PGR to a sound stimulus. The final group of twenty dysthymics were diagnostically constituted as follows: eleven, mainly obsessional and compulsive features; five, mainly anxiety features; four, mainly depressive features. In most cases the diagnoses were mixed, including more than one of these features.

The Personality Questionnaires

Guilford's scales *STDCR* (13) and the Guilford-Martin scales *G* and *A* (16) were typed on cards, one item to each card, and the *S* requested to "mail" each card into one of three slotted boxes labeled YES, NO, and ?. The 237 cards were presented to the *S*s in random order. The seven traits are described in the test manuals as follows:

S—social introversion: shyness, seclusiveness, etc.
T—thinking introversion: an inclination toward meditative or reflective thinking, self analysis, etc.
D—depression: habitually gloomy, pessimistic with guilt feelings, etc.
C—cycloid disposition: strong emotional fluctuations, emotional instability, etc.
R—rhathymia: happy-go-lucky, carefree, lively, impulsive, etc.
G—general activity: a general pressure to engage in overt activity, etc.
A—ascendancy: leadership qualities; social ascendancy etc.

The Guilfords identified these traits by factor analysis studies (14, 15). There is, however, some evidence to suggest that *C* and *D* are largely measures of the factor of neuroticism and that *R* is a good measure of the introversion-extraversion dimension, a high *R* score characterizing the extravert end of the dimension (3, p. 106ff.; 17).

Taylor's Manifest Anxiety Scale and the MMQ were also typed on cards and the *S* requested to "mail" each card into one of two boxes labeled TRUE and FALSE. Taylor's scale normally consists of 50 anxiety items and 175 buffer items, but here the original 175 buffer items were omitted and in their place was substituted the 48 items of the MMQ. These 98 items were administered in random order.

Taylor's anxiety scale has been used extensively in conditioning studies (e.g., 29, 30, 32) and is supposed to discriminate the anxious from the nonanxious *S*s. The MMQ consists of a 40-item neuroticism scale together with 18 "Lie" items. It is discussed in detail by Eysenck (2, p. 94ff.), and there is little doubt that, given good subject motivation, it discriminates well between neurotics and normals.

Conditioning Sequence, Apparatus, and Method of Recording

Before the conditioning session was begun, each *S* was given three tone stimuli, followed by three air puff stimuli (*not* paired with the CS) and then three more tone stimuli. Only *S*s who did not give PGR or eyeblink responses to the last three tone stimuli were included in the conditioning study. The purposes of these trials were to eliminate those *S*s who showed any evidence of pseudo-conditioning (11, 12) or original sensitivity to the tone.

The conditioning session consisted of 30 reinforced trials, randomly interspersed with 18 test trials (called acquisition trials). The US consisted of an air puff, lasting 500 msec., delivered at a pressure of 65 mm. of mercury from a 2.5 mm. internal diameter polythene tube at approximately 2 cm. from the right eye. The CS was a pure tone, delivered to both ears through a pair of padded high quality earphones at a frequency of 1100 cycles per sec. for a duration of 800 msec. after the tone was presented. The time intervals were accurately controlled by means of an electronic timer. Finally, each *S* was given ten consecutive test trials (called extinction trials).

The air was supplied from a compressed air cylinder and puffed into the eye by means of an electronically operated gas valve. The polythene tube ran from the gas valve to a pair of plain glass spectacles worn by the *S*. The right spectacle lens had a small aperture into which the plastic tube was attached. The eyelid movements were recorded by means of a photoelectric cell attached to the same lens as the polythene tube. This photoelectric method of recording the eyelid movements has been described in detail elsewhere (10). It is particularly suitable for working with patients since it does not require any electrodes or artificial eyelashes to be attached directly to *S*, and *S* does not have to keep his head rigidly still. Briefly, it consists of a small photoelectric cell and a linear amplifier, the amplifier EMF being used to drive the pen of a recording milliameter. The milliameter was equipped with several channels so that eyelid movements, PGR trace, and the occurrence of the CS and the US all appeared on the same record.

Procedure

The *S*s were told that they were to be tested in a quiet room where measures could be made of how well they were able to relax under various conditions. As detailed elsewhere, (6) the conditioning laboratory can be partitioned into two by means of a sliding curtain. The *S* was seated comfortably in an armchair in one half of the laboratory, facing a small table so arranged that his field of vision was largely confined to the plain walls of a booth constructed around this table. His head rested against an adjustable padded headrest, and his feet were supported by a footstool. The *E* and the apparatus occupied the other half of the room. Once the test was ready to begin, the curtains could be drawn together, leaving a small gap just sufficient for *E* to see *S* (but *S* was unable to see *E*).

The spectacles, headphones, and PGR electrodes were fitted on *S*; and after a suitable interval, *S*'s hearing threshold was measured for the frequency of 1100 cycles per second (both ears simultaneously). Any *S* whose hearing threshold was below −20 db was excluded from the study. In the middle of the booth was a small red light (6.3 v., 0.5 amp.) which *S* was instructed to look at whenever it was on. The red light was switched on five to ten seconds before any stimulus was delivered and switched off a few seconds afterward. It was found that this procedure reduced strain

in *S* and ensured that his eyes were open during the critical period just before any stimulus was delivered without having to tell him to keep his eyes open. It also reduced eye movements and spontaneous blinking during this critical period. The nine sensitization and pseudo-conditioning stimuli were then given, followed by the conditioning and extinction session.

The *S* was then asked such questions as, "Did you find the air puff bearable or unpleasant?" and "Did you feel sleepy?." Finally, he was requested to sort the personality items. The intelligence test had been given to most *S*s upon an earlier occasion. For the 20 normal *S*s, the whole procedure, except the intelligence testing, was repeated after an interval of 14–21 days.

A CR was recorded whenever the record showed a deflection of 1.27 mm. (.05 in.) or more during a latency of between 156 and 625 msec. after the onset of the CS for the eyeblink and a latency of between 0.5 and 8.0 seconds after the onset of the CS for the PGR.

RESULTS

Table 2 gives the mean number of conditioned responses during the 18 acquisition test trials and the ten extinction test trials for both PGR and eyeblink reflexes in all three groups.

From the data presented in Table 2 it may be concluded for both eyeblink and PGR reflexes that:

1. Dysthymics give significantly more CR's than hysterics both for the acquisition trials and for the extinction trials (one-tailed t test: $p < .0005$ for the eyeblink acquisition and extinction trials; $p < .005$ for the PGR acquisition trials and $p < .025$ for the PGR extinction trials).

2. The normals give significantly less CR than the dysthymics for both acquisition and extinction trials (one-tailed t test: $p < .0025$ for the eyeblink acquisition trials; $p < .01$ for the eyeblink extinction trials; $p < .025$ for the PGR extinction trials).

3. When the normals are compared with the neurotics as a combined group, there are no significant differences in the number of CR's produced.

The differences are not so marked for the PGR as for the eyeblink reflex, probably because of the low intensity of the US. It was noticed that an air puff at a pressure of only 65 mm. of mercury was in several cases insufficient to produce even an unconditioned PGR after the first few trials.

Figure 1 shows the histograms for the acquisition CR's for the eyeblink reflex. If eight CR's are chosen as the cutoff score, then the classification error is 20 per cent for the hysteric group and 10 per cent for the dysthymic group, giving a total misclassification of only 15 per cent.[3] In Fig. 2, the total number of eyeblink CR's given by each of these groups is plotted for each of the 18 acquisition test trials and the ten extinction test trials.

TABLE 2

MEAN NUMBER OF CR'S GIVEN BY EACH GROUP (N = 20 per group; maximum acquisition score = 18, maximum extinction score = 10)

Group	Eyeblink				PGR			
	Acquisition		Extinction		Acquisition		Extinction	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Hysterics	6.7	5.9	2.7	2.7	2.3	2.7	0.8	1.7
Normals	7.9	5.4	3.7	2.7	2.4	3.3	0.6	1.7
Dysthymics	12.4	4.1	6.3	3.2	5.4	4.0	2.1	2.3

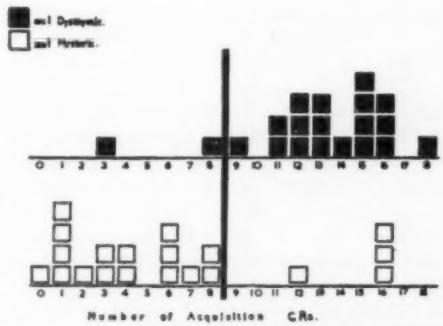

FIG. 1. NUMBER OF HYSTERICS AND DYSTHYMICS MISCLASSIFIED USING THE FREQUENCY OF EYEBLINK CONDITIONED REFLEXES AS A MEASURE

As indicated in Table 2 and Fig. 2, there are no statistically significant differences in conditionability between normals and hysterics. In accordance with the theory presented in this paper, it is to be expected that normals would differ with respect to conditionability from both hysterics and dysthymics to approximately the same extent that they differed from these two neurotic groups in extraversion. In the present study the extraversion scores of the normal group (see Table 1) are more

[3] When similar histograms are plotted for the PGR data using a cutoff of 3 CR's, the total misclassification is 30 per cent. When a double criterion is used, however, in which to be classified as hysteric the patient must have 8 or less eyeblink CR's *and* also 3 or less PGR CR's, then the two groups may be separated with no misclassification whatsoever.

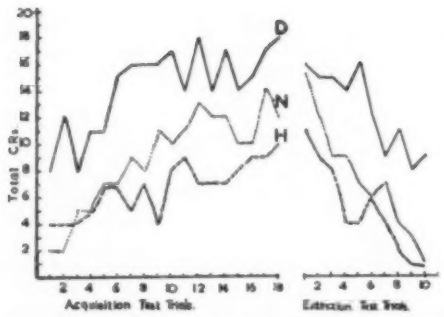

FIG. 2. TOTAL NUMBER OF EYEBLINK CR'S GIVEN BY EACH GROUP AT EACH TEST TRIAL
(N = 20 in each group)

like those of the hysteric group than of the dysthymics. This state of affairs probably occurred because, although the normal group was chosen at random, it was, naturally enough, the extraverted rather than the introverted subjects who volunteered for the experiment. This situation possibly accounts for the findings that the conditioned response behavior of the normals is somewhat similar to that of the hysterics although the differences in conditionability between these two groups are not statistically significant.

Similar results were obtained for PGR conditioning; here also the dysthymics conditioned much more strongly than the hysterics and than the normals, and the hysterics and the normals behaved very much the same. Both the hysterics and the normals seemed to adapt to the PGR fairly rapidly, and the normals tended to resemble the hysterics more than the dysthymics with respect to their extraversion scores.

Another measure of conditionability was the number of reinforcements necessary to achieve a criterion of three successive CR's. This index gave a correlation with the acquisition score of .80 for the eyeblink reflex and .78 for the PGR reflex, which is in agreement with the finding of Humphreys (20), who concluded that to report a criterion score and an acquisition score is superfluous since the correlation between these two scores is as high as the reliability of the conditioning measures will allow.

For the normal group, the test-retest reliability over the period of 14–21 days was .52 for the eyeblink acquisition score, .37 for the eyeblink extinction score and .40 for the PGR acquisition score. The PGR Pearson product-moment coefficient for the extinction data could not meaningfully be calculated because there were too many zero values on retest. The PGR correlation is low when compared with that of Welch and Kubis (36, 37), who obtained a test-retest reliability of .88, but they used electric shock as the US.

When corrected for attenuation (a procedure carried out only where reliability data existed, i.e., for the normal group), the correlation between the PGR and eyeblink acquisition scores was .53. For none of the groups were there any significant correlations between any measure of conditioning and age, sex, or intelligence.

TABLE 3
CORRELATIONS AMONG CONDITIONING AND GUILFORD'S R SCALE, TAYLOR'S ANXIETY SCALE AND THE MMQ FOR ALL Ss
(N = 60)

Scale	Eyeblink Conditioning		PGR Conditioning
	Acquisition	Extinction	Acquisition
R	−.48	−.37	−.25
Anxiety	+.15	+.16	+.17
MMQ	+.08	+.04	+.20

Table 3 presents the Pearson product-moment correlations between conditioning and Guilford's R Scale, Taylor's Manifest Anxiety Scale, and the MMQ for the number of eyeblink CR's during acquisition and extinction. For the PGR only the acquisition CR's were considered, since adaptation proceeded too rapidly during extinction. The correlations between R and the conditioning measures are, of course, negative because rhathymia is a measure of extraversion. Using a one-tailed test of significance, all three correlations between R and the conditioning measures are highly significant, whereas the correlations between conditioning and the MMQ are insignificant in all cases. Similar results were obtained when the three groups were considered separately. Furthermore, for each of the three conditioning measures in Table 3, the correlation between conditioning and R is significantly greater ($p < .02$) than the correlation between conditioning and the Taylor score and the correlation between conditioning and the MMQ.

That the D and C scales are also measures

of neuroticism is suggested by the high correlation (.88) between $D + C$ combined and the MMQ. Taylor's Anxiety Scale correlated .86 with $D + C$ combined and .92 with the MMQ (for all 60 *Ss*). As elaborated elsewhere (8), these intercorrelations strongly suggest that these three scales all essentially are measures of neuroticism; hence, the insignificant correlations of both the anxiety and the MMQ scales with conditioning.

There were several other relevant findings. The hysterics tended to give fewer unconditioned PGR responses to the air puffs than the dysthymics. When the replies to the post-conditioning questions were analyzed, it was found that the dysthymics reported the air puffs as being disturbing significantly more often than did the hysterics. Although not significant, the hysterics appeared to report that they felt sleepy during the session more often than did the dysthymics.

Discussion of Results

These results would indicate very strongly as far as the eyeblink reflex is concerned, and very possibly for the PGR reflex, that conditionability is related to introversion—extraversion and not to neuroticism, the extraverted subjects tending to condition much less well than the introverted ones. The results also suggest that manifest anxiety is related to strong conditionability only to the extent that anxious people are introverted. Taylor's Anxiety Scale differentiates those subjects who condition well from those who condition poorly only to the extent that it fails to measure neuroticism. Since this scale has a very large projection on the neuroticism dimension and only a small projection on the introversion dimension, it is probably an extremely poor diagnostic measure of conditionability. This explains the very low correlation between this scale and conditionability obtained in the present study and also the only slight sucessses obtained by Spence and his colleagues and by Hilgard, Jones, and Kaplan (18) in their attempts to obtain positive correlations between the Taylor scale and conditionability.

The poor conditionability of the hysterics, their more rapid PGR adaptation to the air puffs, their subjective reports that the air puffs were not very disturbing, and perhaps their reports of feeling sleepy all support the hypothesis that hysterics are in a state of cortical inhibition in which dissociation phenomena predominate; in a similar manner, the results obtained for the dysthymic group support the hypothesis that dysthymics are in a state of cortical excitation. Since conditionability in both neurotics and in normals is apparently related to introversion-extraversion and not to neuroticism, it would seem very probable that excitation-inhibition is closely related to introversion-extraversion.[4] This conclusion, if correct, is of considerable theoretical and practical interest, and together with other experimental findings, should help to generate a dynamic and physiologically based theory of personality structure (4) in which the behavior of both the dysthymic and the introverted individual is described in terms of overconditioning and cortical excitation, and the behavior of both the hysteric and the extraverted individual is described in terms of underconditioning and cortical inhibition.

SUMMARY

Pavlov's concepts of excitation and inhibition were related to the dimension of introversion-extraversion in normal and neurotic subjects. Normal and neurotic subjects were conditioned, using the eyeblink and PGR reflexes. It was found that (*a*) anxiety states conditioned much better than hysterics, and (*b*) conditionability is related to introversion-extraversion and not to neuroticism.

REFERENCES

1. EYSENCK, H. J. *Dimensions of personality.* London: Kegan Paul, 1947.
2. EYSENCK, H. J. *The scientific study of personality.* London: Routledge & Kegan Paul, 1952.
3. EYSENCK, H. J. *The structure of human personality.* London: Methuen, 1953.
4. EYSENCK, H. J. A dynamic theory of anxiety and hysteria. *J. ment. Sci.*, 1955, **101**, 28–51.
5. FINESINGER, J. E., SUTHERLAND, G. F., & MCGUIRE, FRANCES F. The positive conditioned salivary reflex in psychoneurotic patients. *Amer. J. Psychiat.*, 1942, **99**, 61–74.
6. FRANKS, C. M. *An experimental study of conditioning as related to mental abnormality.* Unpublished doctor's dissertation, Univer. of London, 1954.
7. FRANKS, C. M. The establishment of a conditioning laboratory for the investigation of personality and cortical functioning. *Nature*, 1955, **175**, 984–985.

[4] This is in agreement with the finding that sodium amytal, an inhibitory drug, decreases conditionability and increases extraversion (9).

8. FRANKS, C. M. The Taylor scale and the dimensional analysis of anxiety. *Rev. Psychol. appl.*, in press.
9. FRANKS, C. M., & LAVERTY, D. G. Sodium amytal and eyelid conditioning. *J. ment. Sci.*, 1955, **101**, 654–663.
10. FRANKS, C. M., & WITHERS, W. C. R. Photoelectric recording of eyelid movements. *Amer. J. Psychol.*, 1955, **68**, 467–471.
11. GRANT, D. A. The pseudo-conditional eyelid response. *J. exp. Psychol.*, 1943, **32**, 139–149.
12. GRETHER, W. F. Pseudo-conditioning without paired stimulation encountered in attempted backward conditioning. *J. comp. Psychol.*, 1938, **25**, 91–96.
13. GUILFORD, J. P. *Inventory of factors STDCR*. Beverly Hills, Calif.: Sheridan Supply Co., 1940.
14. GUILFORD, J. P., & GUILFORD, RUTH B. Personality factors S, E, and M, and their measurement. *J. Psychol.*, 1936, **2**, 109–127.
15. GUILFORD, J. P., & GUILFORD, RUTH B. Personality factors D, R, T, and A. *J. abnorm. soc. Psychol.*, 1939, **34**, 21–36.
16. GUILFORD, J. P., & MARTIN, H. G. *The Guilford-Martin inventory of factors GAMIN, manual of directions and norms*. Beverly Hills, Calif.: Sheridan Supply Co., 1945.
17. HILDEBRAND, H. P. *A factorial study of introversion-extraversion by means of objective tests*. Unpublished doctor's dissertation, Univer. of London, 1953.
18. HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–99.
19. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning*. New York: Appleton-Century-Crofts, 1940.
20. HUMPHREYS, L. G. Measures of strength of conditioned eyelid responses. *J. gen. Psychol.*, 1943, **29**, 101–111.
21. JUNG, J. C. *Psychological types*. London: Kegan Paul, 1924.
22. MCALLISTER, W. R. eyelid conditioning as a function of the CS-US interval. *J. exp. Psychol.*, 1953, **45**, 417–422.
23. MACCORQUODALE, K., & MEEHL, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, **55**, 95–107.
24. PAVLOV, I. P. *Conditioned reflexes*. G. V. Anrep (Trans.). London: Oxford Univer. Press, 1927.
25. PAVLOV, I. P. *Lectures on conditioned reflexes. Vol. 1. The higher nervous activity (behavior) of animals*. W. H. Gantt (Trans.). London: Lawrence & Wishart, 1928.
26. PAVLOV, I. P. *Lectures on conditioned reflexes. Conditioned reflexes and psychiatry*. W. H. Gantt (Trans.). New York: International Publishers, 1941.
27. RAVEN, J. C. R. *Progressive Matrices (1938)*. London: Lewis, 1948.
28. SCHIFF, ETHEL, DOUGAN, CATHERINE, & WELCH, L. The conditioned PGR and the EEG as indicators of anxiety. *J. abnorm. soc. Psychol.*, 1949, **44**, 549–552.
29. SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.
30. SPENCE, K. W., & TAYLOR, JANET A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
31. SPENCE, K. W., & TAYLOR, JANET A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1952, **45**, 265–272.
32. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
33. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
34. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.
35. TAYLOR, JANET A., & SPENCE, K. W. Conditioning level in the behavior disorders. *J. abnorm. soc. Psychol.*, 1954, **49**, 497–503.
36. WELCH, L., & KUBIS, J. The effect of anxiety on the conditioning rate and stability of the PGR. *J. Psychol.*, 1947, **23**, 83–91.
37. WELCH, L., & KUBIS, J. Conditioned PGR (psychogalvanic response) in states of pathological anxiety. *J. nerv. ment. Dis.*, 1947, **105**, 372–381.

Received March 28, 1955.

SELF-APPRAISAL, PERSONAL DESIRABILITY, AND PERCEIVED SOCIAL DESIRABILITY OF PERSONALITY TRAITS[1]

EPHRAIM ROSEN

University of Minnesota

It is widely if not universally accepted that the statements a person makes about his own behavior and personality traits are related to his perceptions of the desirability and acceptability of these behaviors and traits. For example, Edwards (2) found a correlation of +.871 between the probability of endorsement of personality inventory traits and the median-rated social desirability of the traits. The ratings of social desirability were made by a large group of judges who were completely separate from the subjects taking the inventory. Edwards suggests that this high correlation may indicate that desirable traits are in fact widespread in our culture, that subjects consciously or unconsciously try to give good self-impressions, or both. Perhaps a fourth possibility exists: people may believe certain traits are desirable just *because* they are so widespread.

There is thus a problem connected with the basis of this interrelation. A number of other problems, toward which the present research is oriented, also present themselves. For example, can a meaningful distinction be made between two kinds of desirability, which may be called personal desirability and perceived social desirability? Personal desirability is defined here as appraisal of the desirability of given behaviors and traits in oneself, regardless of whether one in fact evidences the behaviors and traits. It is synonymous with "own opinion" of desirability. Perceived social desirability is defined as one's perception of whether or not one's society views the behaviors and traits as desirable, regardless of whether one evidences them and regardless of whether one personally considers them desirable.

If such a distinction can be meaningfully made, what kinds of traits and behaviors are considered personally desirable, and what kinds are perceived as socially desirable? Is there agreement between male and female subjects (*S*s) in such judgments? Is there little or much discrepancy between personal desirability and social desirability? Which of these forms of desirability correlates more strongly with self-endorsement?

The foregoing problems all deal with groups as such. One may also ask a set of clinical, individually centered, questions. For example, what kinds of persons show greater discrepancy between self-endorsement and personal desirability, and what kinds show smaller discrepancies? Between self-endorsement and perception of social desirability? Between personal and social desirability?

Since the Minnesota Multiphasic Personality Inventory (MMPI) was the instrument chosen for this research, it became possible to investigate one other kind of problem, that of the meaning of some of the MMPI scales. Specifically, the meaning of the *K* scale (3) and of the subtle and obvious scales developed by Wiener (5) should be clarified by comparing self-endorsement scores on these scales with personal and social desirability scores on the same scales. In the present report, however, technical discussion of the MMPI will be kept to a minimum, such discussion being reserved for another place.

It should be noted that the MMPI was not designed as a face-valid self-appraisal type of inventory. Nevertheless, in most of the MMPI items the subject makes statements about himself so that for purposes of the present research the use of the terms "self-appraisal" and "self-endorsement" seem justifiable as well as convenient. If an *S* endorses the item, "I enjoy many different kinds of play and recreation," the assumption is no more made here than in clinical use of the inventory that he *actually* likes these activities. It is merely asserted that he has endorsed the item. Therefore, it is inferred that the empirical validation of scales including such items allows conclusions to be made with some probability of accuracy about the personality of subjects, and that comparison can be made between scores on such items and scales under three different conditions of instruction.

[1] This research was made possible by a grant from the Graduate School of The University of Minnesota.

PROCEDURE

The *Ss* were 200 white college sophomores, 102 male and 98 female. All were students in an introductory psychology course, and none had any knowledge of the MMPI beyond the fact that many had taken it routinely on entering the university. They were volunteers and were rewarded for their time by being given extra points in computation of their final course grades.

Each of these *Ss* took the inventory three times. The instructions for the self-appraisal condition were the standard ones on the front of the MMPI group form booklet. For the personal desirability condition, the defining instructions were, "Answer 'true' if you personally think the behavior or characteristic described in each item is more desirable than undesirable in a person of your age and sex; answer 'false' if you personally think it is more undesirable than desirable." For social desirability, the instructions were, "Answer 'true' if you think our society considers the behavior or characteristic described in each item more desirable than undesirable in a person of your age and sex; answer 'false' if you think society considers it more undesirable than desirable." An illustrative hypothetical item was given. Other steps taken to ensure *Ss*' understanding of the task consisted of not giving instructions for the second or third conditions until the previous MMPI form was turned in to the experimenter, and carefully explaining the distinctions between each instruction and the previous ones. It should be noted that the present design differs from that of Edwards in that the same *Ss* make both self judgments and desirability judgments.

To shorten the onerous task of taking the MMPI three times, a form was used containing 409 items instead of the usual 566. These 409 included the first 366 items in the group booklet, which are needed to score the standard clinical scales, and those 43 of the remaining items that are needed to score the *K*, *Si* (social introversion) (1), and *A* (anxiety) (4) scales. With this shortening and considerable encouragement from the experimenter, very few *Ss* who started failed to complete all three forms.

All answer sheets were scored for the following scales: ?, *L*, *F*, *K*, *Hs*, *D*, *Hy*, *Pd*, *Mf*, *Pa*, *Pt*, *Sc*, *Ma*, *Si*, *A*, and the subtle and obvious *D*, *Hy*, *Pd*, *Pa*, and *Ma* scales. In addition, the five scales to which *K* corrections are applicable (*Hs*, *Pd*, *Pt*, *Sc*, and *Ma*) were scored both with and without *K* correction since it was found that applying the *K* correction masked a number of phenomena involving these five scales. The ?, *L*, *F*, *A*, and subtle and obvious scales (*S* and *O*) were kept in raw score form; all others were converted to *T* scores according to the standard norms. Thus, each of the 200 *Ss* was given three scores for each of these scales. For example, each *S* was given a self-appraisal *D* score, consisting of a *T* score calculated from his answers to the *D* items administered under the usual instructions; a personal desirability *D* score, consisting of a *T* score calculated from his answers to the same items under the condition in which he answered true if he thought the item described a desirable characteristic or behavior; and a social desirability *D* score. The latter two scores may thus be interpreted as the amount of depression tendency he personally considers desirable in a person like himself, and the amount he believes society considers desirable in a person like himself. To be sure, many of the *D*-scale items are subtle so that the typical *S* would probably not identify them as depression indicators, but insofar as there is validity in the MMPI scales administered in the usual fashion, it is assumed here that there is likewise validity when administered under the two special conditions. Further, the results to be presented fall into a meaningful pattern which helps support the assumption of validity of procedure.

Half the *Ss* took the three forms in a self-appraisal—personal—social order, and half in a self-appraisal—social—personal order. Comparison of scale means on both the personal and social conditions, separately by sexes and for all scales for which the inventory was scored, showed a few significant differences between the two orders of testing. The number of such differences, however, was no greater than would be expected by chance, and they did not fall into any consistent pattern. It was therefore concluded that there is no evidence that order is an influential variable.

One other extraneous variable was checked for possible effect on the results. Slightly less than half the *Ss* completed all three tasks on the same day, whereas the rest of the *Ss* had an interval ranging between one day and three weeks between starting the first task and completing the third. Analysis of variance showed no scales for which time interval was a significant variable in affecting scores on personal or social desirability.

RESULTS

The first question asked in the introductory section was whether a meaningful distinction can be made between personal and social desirability. One way to answer this question is to examine the correlations between the two kinds of desirability conditions, and also between each condition and the normal self-appraisal condition. For males, the correlations thus obtained ranged from −.02 to +.49, with a median of +.20 for self-appraisal vs. personal desirability; from −.25 to +.43, with a median of +.155 for self vs. social; and from +.03 to +.43, with a median of +.295 for personal vs. social. For females, the three ranges and medians are, respectively: +.07 to +.48, median +.26; −.10 to +.33. median +.105; and +.05 to +.54, median +.33.[2] The correlations are low enough to suggest that the three sets of instructions were genuinely different, as intended. Further, the correlations

[2] The complete table of intercorrelations has been deposited with the American Documentation Institute. Order Document No. 4712 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $1.25 for 35 mm. microfilm or $1.25 for 6 by 8 in. photocopies.

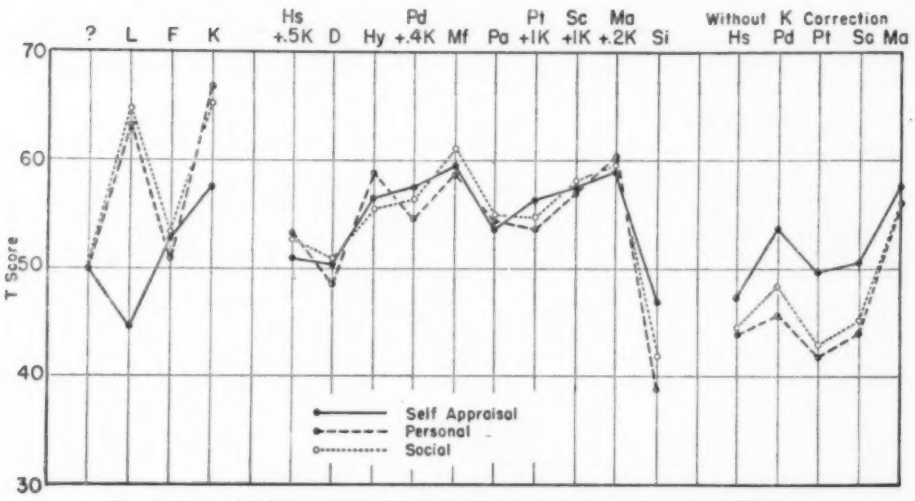

FIG. 1. MEAN MMPI PROFILES (*T* SCORES) FOR THE THREE CONDITIONS: MALES

between personal and social desirability are only slightly larger than between either of these and self-appraisal, so that for the typical *S* in this sample there seems to be discrepancy between personal standards of desirability and perceived social standards. It is also interesting to note that self-appraisal correlates higher with personal desirability for both sexes than it does with social desirability. There is more consonance between perception of own personality and beliefs as to desirable personality than between self-perception and beliefs as to the kind of personality that would conform to social pressures.

These correlations between the three conditions would undoubtedly be higher if the *S*s were as heterogeneous as a clinical population or the normative population upon whom the MMPI was standardized. In the standardization process, scoring was devised in such a way as to yield a standard deviation of 10 *T*-score points in the normal population on each scale. For the present group, median standard deviations of the 16 scales which were converted to *T* scores are 8.75 for self-appraisal, 6.7 for personal, and 7.3 for social, among male *S*s; and in the same order 8.2, 5.4, and 6.65 for female *S*s. Thus the statement that the three conditions constituted three different tests has to be restricted for the present to the college sophomore population. It is interesting to draw certain other conclusions from the median *SD*'s in this group. Males are consistently more heterogeneous than females, so that there is lesser agreement among the men in the group as to own personality, and as to both personal and social standards. To put it otherwise, there is some evidence that women are more similar, and therefore more predictable without individual knowledge of them, than men in the three spheres explored here. Second, both men and women show least variance in personal standards of desirability, more variance in perceived social standards of desirability, and most in self-appraisal of personality. There thus exists, in the present group, a fairly well-accepted norm of personal standards, a less widely accepted perception of social standards, and considerable divergence of own personality.

The second problem to which this research was addressed was that of the kind of traits and behaviors considered desirable on a personal and social basis and the differences between such traits and behaviors in the three conditions. Figures 1 to 4 present mean profiles on all scales for the three conditions. Several conclusions emerge from an analysis of these profiles. First, despite the low intercorrelations of the three forms, except for the *L* and *K* scales there is a basic similarity in shape among the three profiles within both sex groups. There is also a basic similarity between sexes, so that the six mean profiles are variants

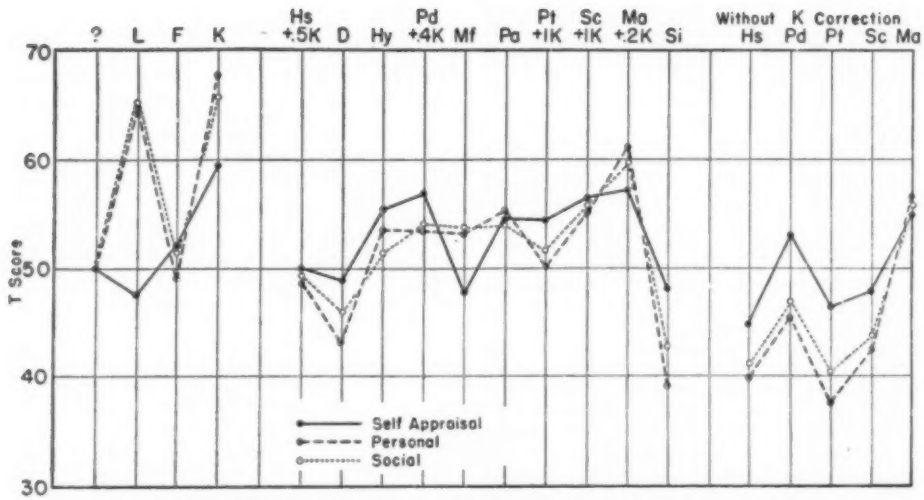

FIG. 2. MEAN MMPI PROFILES (*T* SCORES) FOR THE THREE CONDITIONS: FEMALES

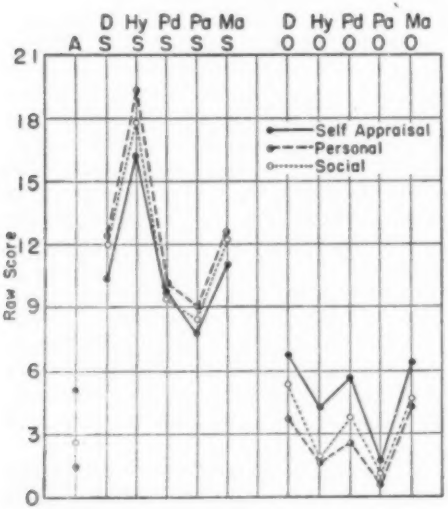

FIG. 3. MEAN MMPI PROFILES (RAW SCORES) FOR THE THREE CONDITIONS: MALES

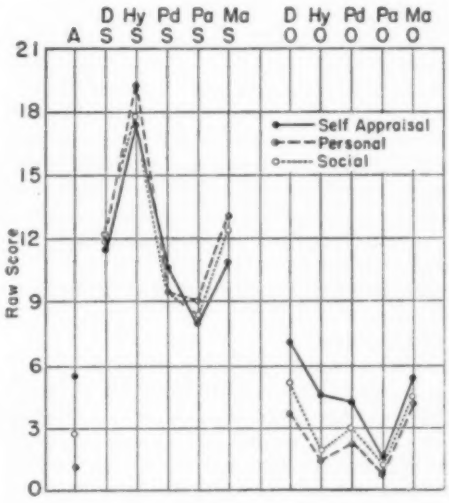

FIG. 4. MEAN MMPI PROFILES (RAW SCORES) FOR THE THREE CONDITIONS: FEMALES

of one configuration, despite elevation differences. To take an example, the average male and female in the group finds *Ma* characteristics more desirable than *D* characteristics, endorses more *Ma* than *D* characteristics, and believes society wants the kind of behavior represented by *Ma* more than it wants the kind of behavior indicated by *D*. This congruence of profiles, coupled with the low intercorrelations, points to the following inference: though most individuals showed but little linear relation between self-appraisal and the two types of desirability, the group as such emerges with an average self-appraisal, an average personal preference, and an average perceived social norm which are highly consistent with each other. There is a perceived group norm *across* personality scales which is

TABLE 1

SIGNIFICANT DIFFERENCES BETWEEN MEANS IN THE THREE CONDITIONS

Male (N = 102)

Self vs. Personal		Self vs. Social		Personal vs. Social	
+	−	+	−	+	−
F(*)	*L*(17)	*Hs*(**)	*L*(14)	*Hy*(**)	*F*(*)
Hs(8)	*K*(10)	*Pd*(5)	*K*(7)	*Hy-S*(**)	*D*(**)
Pd(9)	*Hs+K*(*)	*Pt*(8)	*D−S*(**)	*Pa−S*(**)	*Pd*(**)
Pd+K(*)	*D−S*(6)	*Sc*(6)	*Hy−S*(**)		*Pd+K*(*)
Pt(10)	*Hy−S*(8)	*Si*(**)	*Pa−S*(**)		*Mf*(*)
Pt+K(*)	*Pa−S*(5)	*A*(5)	*Ma−S*(**)		*Pt*(*)
Sc(8)	*Ma−S*(6)	*D−O*(**)			*Sc*(**)
Si(8)		*Hy−O*(6)			*Si*(**)
A(9)		*Pd−O*(5)			*D−O*(5)
D−O(6)		*Ma−O*(5)			*Pd−O*(**)
Hy−O(7)					*Pa−O*(**)
Pd−O(8)					
Pa−O(**)					
Ma−O(7)					

Female (N = 98)

F(**)	*L*(15)	*Hs*(**)	*L*(12)	*K*(**)	*F*(**)
Hs(7)	*K*(9)	*D*(*)	*K*(5)	*Hy*(**)	*Hs*(**)
D(5)	*Mf*(5)	*Hy*(**)	*Mf*(5)	*Hy−S*(**)	*Pd*(**)
Hy(**)	*Ma+K*(**)	*Pd*(6)	*Ma−S*(**)	*Pa−S*(**)	*Pt*(**)
Pd(8)	*Hy−S*(5)	*Pd+K*(**)		*Ma−S*(**)	*Pt+K*(*)
Pd+K(**)	*Pa−S*(**)	*Pt*(6)			*Sc*(**)
Pt(12)	*Ma−S*(6)	*Pt+K*(**)			*Si*(7)
Pt+K(5)		*Sc*(**)			*A*(**)
Sc(8)		*Si*(5)			*D−O*(5)
Si(9)		*A*(**)			*Hy−O*(*)
A(10)		*D−O*(**)			*Pd−O*(**)
D−O(7)		*Hy−O*(6)			*Pa−O*(**)
Hy−O(8)		*Pd−S*(**)			
Pd−S(**)		*Pd−O*(**)			
Pd−O(7)		*Ma−O*(*)			
Pa−O(5)					
Ma−O(**)					

Note.—The 5% level of significance is denoted by (*); the 1% level by (**); and the actual size of all t's of 5 or greater are reported in parentheses, to the next lower integer. The plus sign indicates that the mean of the first of the two variables at the head of the column was larger; the minus sign indicates that the mean of the first variable was smaller. Thus for male, *L*, self vs. personal, the *L* score was lower in self than personal, with a value of over 17 for t calculated for significance of the difference between correlated means.

congruent with both group desirability judgments and mean personality scale scores, even though there is little individual predictability *within* a scale from perceived norm to personal preference to own behavior, or vice versa.

Second, there are elevation differences. Significant differences between mean scores in the three conditions are summarized in Table 1. Many of these differences are very highly significant. Examination of the table and of Figs. 1 to 4 shows that the greatest elevation differences occur in the *L* and *K* scales, a number of the uncorrected clinical scales, the *Si* scale, *A* scale, and four of the obvious scales, when self-appraisal is compared with either personal or social desirability. The average individual would thus like to be, and thinks society wants him to be, more "perfect," even when such perfection has almost a zero actual frequency (*L* scale), more defensive and controlled, less psychopathic (perhaps "impulsive" would be a better term for the population from which this sample is drawn), less anxious and compulsive, less fantasy-and bizarre-minded, considerably more extroverted, and evidencing less of the obvious sort of depressive, hysteroid, psychopathic, and hypomanic kinds of manifestations. Yet these desirability discrepancies from perceived actuality occur in the absence of any marked pathology in actuality. The

average member of the group is relatively defensive, nonpsychopathic, nonanxious, etc. to start with, but he considers it desirable to be even more so.

Discrepancies between self-appraisal and either of the two desirability forms are not solely a function of absolute amount of the trait acknowledged on self-assessment, because to some extent there are distinguishable male and female desirability norms. The *Mf* comparisons are interesting in this connection: males show some elevation in feminine interests, feel that society wants them to show even more of these interests, probably as a function of their adopted role as liberal arts college students, but personally find somewhat greater masculinity of interests desirable. Their conflict is thus essentially a personal-social one. Females, on the other hand, have a quite feminine mean score and both personally and socially find it desirable to have more masculine interests, again probably as a function of their role as college students, for it is well known that college students progressively show more and more interests which in the normal population are associated with the opposite sex. But they personally consider this role expectancy a desirable one, whereas the average male feels he is pulled in a direction opposite to his own standards.

A comparison of self-personal differences with self-social differences reveals that the former tend to be larger, even though self-personal correlations were on the average somewhat larger than self-social. All significant differences between self and social are repeated between self and personal, whereas the converse is not true. In general, one may say, since the self profiles are less "normal" than the desirability profiles, that the subjects' own standards are less accepting than they think society is of whatever disturbances they see in themselves.

The personal-social differences are small but not absent. Social scores are significantly higher than personal for many of the clinical scales, for *Si*, *A* (in females), and some of the obvious scales, as if it was felt that society either expects or actually demands some worry over health, some impulsive behavior, compulsive behavior, anxiety, etc. But social is significantly lower than personal on *K* (in females), *Hy*, and some of the subtle scales which are correlated with *K* (5). There is some evidence here that at least in females the defensiveness measured by *K* is not merely a function of direct social pressures but, over and above these, of individual standards. (Of course, personal standards may be considered indirectly social in the sense that they can be conceptualized as individual introjections of social pressures.)

Up to this point all analysis has been in terms of the scales of the MMPI. But it is also possible to take the 409 items and compare the three forms in terms of the number of items answered true on one form and false on another by each *S*, regardless of the scales in which the items fall. For certain purposes this is a simpler procedure than scale comparison and also involves fewer assumptions. This procedure was followed for the self-appraisal and personal desirability conditions only. A count was first made of the number of items for each subject answered one way in self and another way in personal. The number of ? responses was very small and may be ignored in the ensuing analysis. It was found that for males the range of such discrepantly answered items was from 15 to 165, whereas for females it was 35 to 165, or from about 4 to 40 per cent for the total group. The medians were 85.0 and 87.5, respectively, and the standard deviations were 33.9 and 25.0. Thus, evidence of differences in individuals between self-appraisal and desirability again emerges. Further, the same trend toward greater male heterogeneity, this time in amount of discrepancy, is apparent.

An examination was then made of those specific items most frequently answered "true" in the self or personal condition and "false" in the other. Fifty-six items were answered discrepantly by 30 or more males, and 42 by 30 or more females. Twenty-nine of these were common to both sexes, so that the males tended to be discrepant on most female discrepant items and to have another group of discrepant items in addition. There were, however, no reversals in direction: all items more frequently answered "true" by men in personal than in social, for example, were answered similarly by women, but often below the arbitrary cutting point of 30. The items appear to fall in four classes. First, a subgroup of several *L* items, in which subjects "lie" on the personal form but not on the self form, was identifiable among the 56 highly discrepant items. Second, the largest subgroup consisted of items describing widespread faults in oneself; these

were answered "true" in the self-appraisal condition and "false" in personal desirability, or reversed if the item asserted the absence of such faults. Examples of such items are, "I have difficulty starting to do things," and, for the reversed situations, "I daydream very little." Third, there were several items that constitute negative comments about other people; for example, "I think nearly anyone would tell a lie to keep out of trouble." Such items were answered "true" more often in self-appraisal than in personal desirability, but it is not clear for such an item if a negative answer in the desirability condition means that *S* believes he should not think this (the meaning intended by the instructions), or that people should not lie to keep out of trouble, or that people should not lie. The instructions may have been misunderstood for such items, but fortunately they are relatively few in the MMPI. Last, there were items expressing interests such as, "I like dramatics." These were answered "false" on self, "true" on personal desirability. It was noteworthy that more female *S*s were discrepant on the interest items than male *S*s, and, most strikingly, that the direction of discrepancy was identical whether the interest expressed was a so-called feminine or a so-called masculine interest. Thus female *S*s answered true on personal more often than on self-appraisal for "I like mechanics magazines" as well as for "I would like to be a nurse." They do not appear to feel they have the wrong kind of interests, but rather that they have too few interests of any kind.

Items showing large differences in the number of discrepancies, though never in direction, between men and women tended to be either overt behavioral items of the type, "In school I was sometimes sent to the principal for cutting up," or items of the type, "Once in a while I think of things too bad to talk about." This last was discrepant for 34 males but for only 12 females; either females admit such thoughts less frequently, or, less likely, fewer of them consider such thoughts undesirable.

The discrepancy analysis was used in still another way as a test of Edwards' findings. For males and females separately, the number of *S*s who answered "true" to each of the 409 items in the self-appraisal condition was correlated with the number who answered "true" to each item in the personal condition. It should be noted that even though this approach differs from that of Edwards in that the same group of *S*s who endorse items also make desirability judgments, nevertheless an item can be answered "true" by the same number of people in both conditions yet not by the same individuals; a discrepancy in one direction can be balanced by an equal discrepancy in the opposite direction. The correlations were identical with those reported by Edwards; for both males and females it was +.87.

In an attempt to help choose between the several possible reasons for this relationship, the number of discrepancies was correlated with the *K* scale on the self-appraisal form. The reasoning behind this procedure was that if *K* can be taken as a measure of defensiveness, then a high correlation can be taken as evidence for the hypothesis that the previously reported correlation—that between self-endorsement and desirability—is at least in part due to the conscious or unconscious attempt to give good self-impressions. The correlation between *K* scores and discrepancies was high enough to lend support to this hypothesis. For males it was −.60 and for females −.65, both strongly significant statistically. Thus, it may be said that the degree of defensiveness defined by *K* is a fairly good predictor, in this sample, of the tendency *not* to call traits and behaviors undesirable if they are believed to be present in oneself, and not to call them desirable if they are believed absent. Of course, it is not demonstrated that the perception of desirability *causes* self-endorsement; recognition of the presence of a characteristic in oneself may cause the perception of desirability. Either of these lines of reasoning, however, is consistent with an interpretation putting weight on unconscious or conscious defensiveness.

There are other types of analysis of the data which might shed further light on the nature of the role of desirability in endorsement. Correlation with the *Hy S* scale might be valuable, as might correlation with such scales as *Sc* and *Si*. In addition, an analysis of the relation of each individual's discrepancy score to the kind of profile he shows in self-appraisal should be useful though laborious. Investigation could also be made into the question of what sorts of individuals show large discrepancies, contrary to the group trends, between personal and social desirability.

In addition, further work can profitably repeat the entire study on those subgroups of clinical patients who can follow the instructions. Some of the present results would probably be unchanged in such a group and some would undoubtedly be quite different. For example, it can be predicted that there would be greater elevation differences between self-appraisal and personal desirability scale scores, and probably, for at least some clinical subgroups, between personal and social desirability.

SUMMARY

An investigation was made of a number of problems in the interrelations among endorsements of statements about one's own personality traits and behavior, one's personal appraisal of the desirability of these traits and behaviors, and their perceived social desirability. The instrument used was the MMPI. Two hundred college sophomores took the MMPI three times: in the usual manner (self-appraisal), answering "true" if the item described a behavior or trait which was personally considered desirable in a person of one's own age or sex (personal), and answering "true" if it was thought that society would consider it desirable in a person of one's own age and sex (social). The three forms were scored for each *S* on the validity and clinical scales and on the following additional scales: *Si*, *A*, and the subtle and obvious scales. In addition, the number of individual items on which each *S* answered discrepantly between the self-appraisal condition and the personal condition was counted. Analysis showed the following results:

1. Variation of order of testing in the personal and social conditions seemed to make no difference. Time between beginning and completing the three tasks made no difference.

2. Intercorrelations between the three forms were low enough to show that the three tasks were genuinely different. The concept of desirability may be subdivided into personal desirability and perceived social desirability. For the group tested, self-appraisal correlated higher with personal than with social desirability.

3. Variance on all three forms was lower than in the standardization of the MMPI. The greatest homogeneity appeared on personal and the least on self-appraisal for both men and women. Women were more homogeneous than men.

4. The mean profiles were highly similar in all three groups so that the group norms of personal and social desirability were similar to actual distributions of personality characteristics.

5. There were differences in elevation, however. The clinical scales were in general highest on self, as were the *Si*, *A*, and obvious scales. The *L*, *K*, and subtle scales were, in general, lowest on self. Thus even the small amount of psychopathology present in the group was considered undesirable, and defensiveness and extraversion were considered desirable. Self-personal differences tended to be larger than self-social despite the fact that self-personal intercorrelations were somewhat larger.

6. Male and female profiles were similar but some interesting differences appeared, particularly in the *Mf* scale. There was evidence that masculinity-femininity of interests was a source of conflict for both sexes but in different ways for the two.

7. Results based on analysis of item discrepancies rather than scale comparisons showed that discrepancies tended to be very sizable, that the items most likely to be answered discrepantly fell in a few meaningful classes, that the correlation of endorsement with rating as desirable by items was +.87, and that the number of discrepant items by subjects correlated very strongly and negatively with the *K* scale. This last result was interpreted as throwing some light on the sources of the importance of desirability in endorsement of items.

REFERENCES

1. DRAKE, L. E. A social I.E. scale for the MMPI. *J. appl. Psychol.*, 1946, **30**, 51–54.
2. EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.
3. MEEHL, P. E., & HATHAWAY, S. R. The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, **30**, 525–564.
4. WELSH, G. S. Factor dimensions of the MMPI University of North Carolina, 1954. (mimeographed.)
5. WIENER, D. N. Subtle and obvious keys for the MMPI. *J. consult. Psychol.*, 1948, **12**, 164–170.

Received February 11, 1955.

DISCRIMINATION AND RELIABILITY IN Q-SORT PERSONALITY DESCRIPTIONS[1]

NORMAN H. LIVSON AND THOMAS F. NICHOLS

Guidance Study, Institute of Child Welfare, University of California

In their review of some of the considerable recent personality research employing the Q sort[2], Seeman and Raskin note that "the Q-sorting instrument. . .provides an efficient method of securing a large number of ratings which can be compared from person to person" (8, p. 223). Although the Q sort, as conceived by Stephenson (9), was intended to serve as a tool within the broader framework of his Q methodology, it has proved readily compatible with the demands of a wide variety of research problems in psychology. The fact that so essentially simple a technique as the Q sort should possess such a particular congeniality with so broad a range of personality problems may perhaps be attributed to its ability to speak both the language of the clinician's highly idiosyncratic case description and the researcher's quantified generalizations. It facilitates communication between these two "attitudes," a condition crucial to effective research in the area of personality.

Perhaps the property of the Q sort most responsible for its communicative advantages in personality description is its ipsative nature: the individual being described provides his own frame of reference. No normative judgments as to the amount of a given characteristic relative to some reference group are required but, instead, items are arranged in order of their degree of "importance" in characterizing the person. This dimension of "importance" is likely to be a highly personal one: the sorter may employ some highly complex, multidimensional, perhaps even unverbalizable, criterion for the ordering of the Q-sort items. Furthermore, any given item must be evaluated in the context of all other items. An item may be sorted as highly "important" although it does not occur with unusual frequency, intensity, or duration in the person. Thus, the Q sort allows for the quantitative expression of the feelings and "hunches" of clinical observation. We venture that it is this feature of the Q sort which underlies its success in conjoining the efforts of treatment and research workers in the field of personality.

There has been considerable controversy with respect to the specific form which Q-sort response distributions should take. The issues involved may be summed up in two questions. (*a*) Should the sorter be allowed to distribute the items as he chooses among the available categories of response (unforced Q sort); or should he be required to conform to a prescribed frequency distribution (forced Q sort)? (*b*) If the sorter is to be forced, what specific frequency distribution should be employed for the forced sort?

With regard to the first question, Cronbach has recently supported the use of the unforced sort (2, 3). He is concerned with the possible loss of useful information as a result of the elimination of elevation and scatter in the forced sort. Pitted against this possible disadvantage is the fact that the forced sort insures statistical intersorter comparability of data (since all sortings conform to the same distribution and thus have the same mean and the same variance) as well as thereby facilitating usual computational operations. The first step, then, is to assess the utility of the additional information contained in the unforced sort.

Block (1), using the same Q-sort items and social objects, sought to determine the nature

[1] The authors wish to acknowledge the collaboration of Dr. Jack Block in the data collection and express their appreciation of many fruitful discussions with him during the preparation of this report. We also wish to thank Dr. Harold E. Jones, Director of the Institute of Child Welfare, for his critical reading of the manuscript, and Dr. Jean W. Macfarlane, Director of the Guidance Study, under whose auspices this study was conducted.

[2] Very broadly speaking, the Q sort is merely a forced-choice rating procedure. As used in personality description, it usually consists of the assignment of a large number of descriptive statements to a fixed number of ordered categories. The extreme categories contain those statements judged to be "most characteristic" and "least characteristic" of the person being described. The number of statements which may be placed in each of the categories is usually prescribed. For a succinct description of Q sorting see Stephenson (9, p. 115).

of this additional information by investigating intersorter variation in unforced Q sorts. By means of a double-classification analysis of variance, he has shown that variance in elevation and scatter reflects intersorter response idiosyncracy, but makes little, if any, contribution to discrimination among objects. Therefore, since in this report we are concerned with reliable descriptions of social objects, we conclude, in answer to the first question, that a forced distribution is preferable in our situation.

Turning to the second question, the distribution for the forced sort, we note that, to date, there has been an almost exclusive use of a quasi-normal distribution for the forced Q sort. The reasons for this are not clear. Stephenson (9) has recommended the use of such a flattened symmetrical distribution but it should be recognized that the statistical requirements of certain of his methods of analyzing Q-sort data underlie this recommendation. Contrariwise, the comparisons of Q-sort descriptions by correlational or by item analyses usually impose no such statistical requirements. Another factor which may have contributed to the wide use of the quasi-normal distribution is the assumption that the "natural" distribution of attributes tends to take this form. Ferguson (4), in a different context, cogently discusses this point. He observes that "the vague belief exists that certain abilities are normally distributed in the population, a statement which is meaningless unless ability is given an operational definition. . . . The assumption of normality in many arguments rests on a philosophical basis only and should not be invoked unless it can be shown to be both necessary and useful in a particular context" (4, p. 68). Since the actual preference of sorters appears to be the only means for operationally defining the population distribution of personality characteristics, this question can be answered empirically rather than decided by fiat.

This study will deal with two problems. (*a*) Is there a preferred distribution as indicated by the free choices of distributions in the unforced Q sort? If so, what is its form? If there is neither intersorter communality nor intrasorter stability in these choices, it would be reasonable to assume that no particular distribution, including the quasi-normal one, deserves special status as the preferred "natural" distribution. (*b*) What would be the most efficient form for the forced Q sort? Amount and reliability of information carried by a Q sort will be the efficiency criterion.

METHOD

Social objects. We chose as our objects Harry Truman (HT), Dwight Eisenhower (DE), Joseph McCarthy (JM), Robert Taft (RT), and Adlai Stevenson (AS). While the personality assessment of political figures may seem rather far-removed from the usual clinical use of Q-sort procedure, we found it most appropriate for our purposes. Since a relatively standard pool of data is released concerning such persons, it could be assumed that about the same amount and kind of information was available concerning the objects.

Q-sort items. A set of 76 items, developed at the Institute for Personality Assessment and Research, was used in this study. These items were designed to permit comprehensive and detailed personality descriptions. The language employed is relatively nontechnical and caused the sorters little difficulty in using these items to describe the social objects. Each item was typed on a 3 by 5 card and the 76-item deck was shuffled before each new sort.

To illustrate the scope of this item set, we may cite some of the classes of items with an example of each: interpersonal ("is cold and distant in his relations with others"); social stimulus value ("tends to arouse hostility and resentment in others"); inferential ("lacks insight in his own motives and behavior"); characterological ("is a conscientious, responsible, dependable person"); surface descriptive ("has slow personal tempo; responds, speaks, and moves slowly"); predictive ("would become confused, disorganized, and unadaptive under stress"). Almost all the items were considered as evaluative by the sorters although the set was so designed as to provide approximately equal numbers of generally "good" and "bad" characteristics.

Sorters. Eleven persons (six females, five males) served as sorters in this study. Of these, six were graduate students in psychology, two were college-trained social workers, two were secretaries in a psychological institute, and one was a housewife. While several sorters had some knowledge of the Q-sort technique, none had any previous experience as sorters. While our sorters did differ in the amount of information they had about the objects, all sorters were sufficiently informed to the extent that they did have some "picture" of each object.

Procedure. Each sorter judged the five objects in the same order (HT, DE, JM, RT, AS). Starting with the unforced sort, the sorters were instructed to assign each of the 76 items to one of nine categories of judgment. These categories ranged from "extremely characteristic" of the object (Category 1) to "extremely uncharacteristic" of the object (Category 9). The only restriction which was imposed on the sorters was that they use no more than nine categories. Beyond this, the sorters were left to choose the distribution most comfortable for them. They were not required to use all the categories and were free to assign as many items as they wished to any category. Some of the sorters

made a preliminary sorting of the items into three piles in order to facilitate the sorting.

When the unforced sort had been completed and the data were recorded, the sorter was then instructed to rearrange the items into a quasi-normal distribution (three 1's and 9's, five 2's and 8's, nine 3's and 7's, thirteen 4's and 6's, and sixteen 5's). This rearrangement was done in the following way: working from the extremes toward the middle, excess items were shifted and vacancies were filled from immediately adjacent categories. It was often necessary to move a given item over two or more categories in order to bring the unforced distribution in conformity with the forced distribution. Forced item placements were then recorded. An average of 20 minutes was required to complete both sortings, about five minutes of this time being spent in shifting the unforced to the forced distribution.

After the sorters had completed the two sortings of each of the five objects, each sorter was asked to rank the five objects along two dimensions: (*a*) information they felt they possessed about the object; (*b*) ease and confidence of judgment. This completed Run I.

Due to differences in the availability of sorters, the time intervening between Run I and Run II varied from only a few days for some sorters, to two months in the case of others. In view of the large number of responses involved in making the ten Q sorts (5 forced, 5 unforced) it seems very unlikely, even for the least intervening time, that immediate memory of previous item placements played a role in the judgments made in the second run. The procedure in Run II was identical with than in Run I with the exception that the rankings of the objects were not repeated.

The above procedure provided data from which the following measures were derived: (*a*) test-retest reliabilities for both unforced and forced sorts, and (*b*) discrimination scores, i.e., the number of interitem discriminations inherent in a given distribution of item placements computed by Ferguson's (4) method.

RESULTS AND DISCUSSION

Is there a preferred distribution? Our procedure yielded a total of 110 Q sorts obtained in the unforced condition in Runs I and II. Table 1 provides a summary description of the use of the nine available judgment categories for the unforced placement of the 76 Q-sort items by our eleven sorters. The magnitudes of the intracategory sigmas and ranges indicate a striking variability in the use of the nine categories. Although the individual unforced distributions are almost always symmetrical about the middle category and usually unimodal, there is a marked heterogeneity of shape, ranging from near-rectangularity to extreme leptokurticity.

It might be argued that, although there is such heterogeneity of shape present in unforced sorts, the quasi-normal form of distribution would still be closest to the average preference. But the composite distribution, which can be read from the means of Table 1, is clearly not quasi-normal; by chi-square test it differs significantly from the distribution employed for the forced sorts ($p = .01$). The individual object composite sorts are likewise nonnormal in shape. In fact, the shape of the over-all composite distribution, excluding the middle category, is rectangular: the departure from perfect rectangularity is not significant ($\chi^2 = 6.57$, $.50 < p < .70$). If closeness to the average were the sole criterion for the optimum shape of the forced distribution, a perfectly rectangular shape would be a possible choice. It is evident that there is little intersorter agreement on the quasi-normal or on any other shape as the most "natural" or preferred for the unforced sort in our data.

TABLE 1

CHARACTERISTICS OF PLACEMENT OF 76 ITEMS IN 110 UNFORCED Q SORTS

Statistic	Category								
	1	2	3	4	5	6	7	8	9
Mean	6.24	7.70	9.69	7.97	14.96	6.91	6.67	7.84	8.14
SD	4.35	4.87	5.00	6.45	7.35	6.42	4.79	4.59	6.16
Range	0–19	0–22	0–24	0–27	0–34	0–29	0–25	0–19	0–29
Frequency in Forced Sorts	(3)	(5)	(9)	(13)	(16)	(13)	(9)	(5)	(3)

The point might still be made that, despite an absence of a stable intersorter preference, any individual sorter does have his own preferred distribution. We can investigate this possibility. The discrimination score, in the case of unimodal distributions, is a monotonic function of the shape of the distribution, increasing from leptokurticity to rectangularity. Inasmuch as the sorters' unforced distributions were nearly always unimodal, we used their discrimination scores as a rough index of the shapes of their unforced distributions. The correlation between the Run I and Run II discrimination scores (based upon 55 paired sorts) over the entire group was only .39. While this is a statistically significant relationship, it points to a very low degree of resemblance between the two sets of Q sorts. It appears, then, that neither intrasorter stability nor intersorter communality of distribution preference is sufficient to recommend any one distribution as the preferred "natural" distribution for sorters, either individual or group.

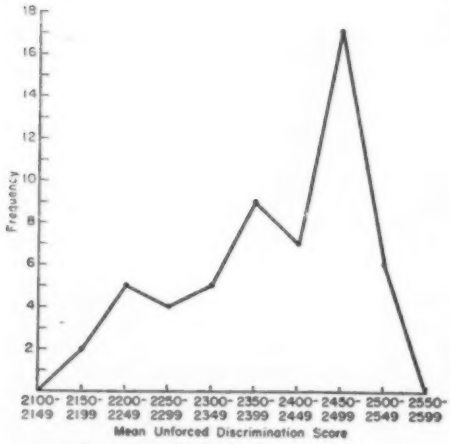

FIG. 1. FREQUENCY DISTRIBUTION OF MEAN DISCRIMINATION SCORES FOR UNFORCED Q SORTS (N = 55)

The median is 2416, the mean is 2395. The quasi-normal distribution used for the forced sort involved 2476 discriminations. The maximum possible discrimination score which could be made among 76 items with nine categories is 2566 (a rectangular distribution).

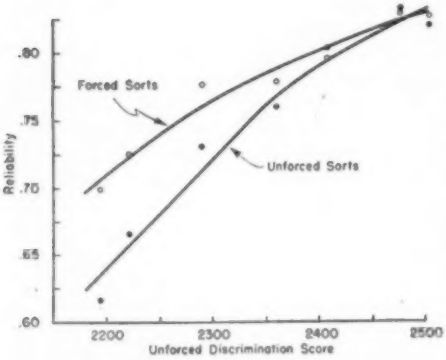

FIG. 2. TEST-RETEST RELIABILITY OF FORCED AND UNFORCED SORTS AS A FUNCTION OF UNFORCED DISCRIMINATION SCORE

Each curve is based upon 55 reliability coefficients. The plotted points were obtained from grouped data by a smoothing process; the end points determined arithmetically and the remainder graphically. Each point is based upon at least seven observed values. The curve is freehand.

What distribution is most efficient for the forced Q sort? We can now go on to evaluate the relative efficiencies of different distributions. As previously stated, the optimum distribution will be the one which permits the sorter to provide maximum information (inter-item discriminations) about the object without sacrificing reliability. This is not to assume that increasing reliable discriminations necessarily increases the validity of the information obtained. In the absence of any independent criterion for validity, however, the maximization of reliable information must become our criterion.

Figure 1 presents the distribution of the mean discrimination scores. The range of these data is sufficient to permit a valid estimate of the relationship between the number of discriminations made in an unforced sort and the reliability of the unforced sort. For each sorter, the mean discrimination score of his two unforced sorts of each object was computed, and related to his unforced reliability. Parenthetically, it should be noted that although there was considerable variation in the test-retest time interval, no relationship between interval and reliability (forced and unforced) was found. From the lower curve in Figure 2, it is clear that, as more discriminations are made, the total pool of discriminations becomes more reproducible.

Of course, this relationship may be mediated by a third factor, namely, the degree of difficulty of the particular sort for a given sorter. Merely from the assumption that sorts which are less difficult to make permit more discriminations and higher reliability, it would follow that discrimination score and reliability would be positively related. Since sorters undoubtedly possessed different amounts of information about the several objects upon which to base their discriminations, the sorts varied in difficulty for each sorter. We can make an estimate of the difficulty of a given unforced sort for a given sorter by regarding the reliability of the corresponding forced sort as a measure of the difficulty of making reliable discriminations in that sort. These estimates have intersort comparability since they were all made at 2,476 discriminations. First, we can use this measure to determine whether, in fact, more discriminations were made in less difficult sorts. The upper curve in Figure 2 indicates that this is so: lower forced reliability coefficients tend to be associated with fewer discriminations. In order to express the extent of the two relationships depicted in Figure 2, we

converted the reliability coefficients to Z scores and computed product-moment correlations. Discrimination score correlates .47 ($p < .001$) with unforced reliability and .36 ($p < .01$) with forced reliability. Carrying this analysis further, we can partial out forced reliability (which measures sort difficulty) from the correlation between discrimination score and unforced reliability. Doing this, we find a significant positive relationship remaining ($r = .33$; $p < .02$). This finding implies that, regardless of sort difficulty, making more discriminations tends to increase the reliability of the total information carried by the sort.

Figure 3 illustrates this increase in reliability as a function of increased discrimination with forced reliability held constant: the curve is essentially the smoothed regression line for the above partial correlation coefficient. Regardless of sort difficulty, a sorter who chooses to make fewer than average discriminations sacrifices reliability; if he makes more than an average number of discriminations, he tends to increase, or at least suffers no loss in reliability. It appears that the particular discriminations which a sorter is reluctant to make can be, in fact, made quite reliably. Since the median unforced discrimination score is greater than 2,400, we can read from the curve that the average could be forced to 2,500 discriminations with no loss in reliability.

We can go one step further and suggest that, even if all sorters had been forced to use a rectangular distribution which would require the maximum possible number of interitem discriminations (2,566 among the 76 items), no appreciable loss in reliability would have resulted. Within the generalizability of our data, therefore, we conclude that sorters may be required to employ a forced rectangular distribution which requires them to make the maximum number of interitem discriminations possible in the Q-sort set and still provide reliable information.

In making this recommendation we realize that we have not as yet considered the possible effects of our violation of the oft-repeated dictum that the sorter should feel "comfortable" about the sorting procedure. For example, Stephenson says, "It is important that the operation itself (Q sorting) should be a reasonable one, such as a person can perform without feeling that it gravely distorts what he wishes

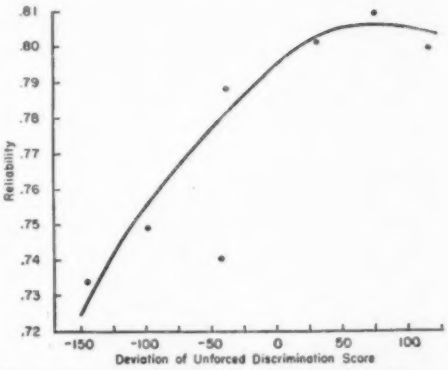

FIG. 3. RELIABILITY OF UNFORCED SORTS AS A FUNCTION OF UNFORCED DISCRIMINATION SCORE INDEPENDENT OF SORT DIFFICULTY

The abscissa is the difference between the observed unforced discrimination score and the mean discrimination score for all unforced sorts falling within the given forced reliability class interval. The plotted points were obtained from 55 unforced reliability coefficients by a smoothing procedure. Each point is based upon at least 13 observed values. The curve is freehand.

to do" (9, p. 60). The point, now, is to investigate how the sorter felt about his task and determine the relation of those feelings to his sorting efficiency.

Some observations on sorter phenomenology. Our sorters appeared to understand readily the procedure of Q sorting. Since none of the sorters had ever Q sorted before, and since about half of them had had no previous acquaintance whatever with the technique, such evidence points to the suitability of the Q sort as a method for the expression of judgment by untrained persons. Although acceptance of the procedure was generally high for the unforced sort, it became somewhat strained under the forcing condition. As might be expected, the more items a sorter had to move, as a result of forcing, the greater the amount of discomfort expressed. By far the most frequent objection was to the insufficient number of items allowed in the extreme categories. The middle category was also singled out for its restricting effect. In this case, however, the required frequency was as often felt to be too high as too low. Summarizing, the sorters felt that an element of randomness often entered into their decisions concerning the movement of items in and out of categories to meet the frequency requirement of the forced normal shape and

believed that this factor probably attenuated the reproducibility of their judgments. Incidentally, it is encouraging that neither in these reactions nor in the statistical characteristics of the Q-sort productions did we find any systematic relationships between the education and training of the sorter and his experimental behavior.

It is unlikely that the sorters' experienced "distortion" of judgments in the forced condition must be taken very seriously since, as we pointed out in our earlier discussion of unforced distribution preferences, the unforced distribution itself has very low reproducibility. In fact, the considerable change in the unforced distribution, for a given sort, from Run I to Run II would, in itself, tend to attenuate reliability even if the judgmental ordering of the statements remained the same. Furthermore, since we have shown that reliability would be only slightly reduced even if maximum discrimination were to be enforced, forcing could hardly be expected to affect reliability seriously.

Perhaps the most frequent and potentially serious reaction of our sorters was their insistence that they had insufficient information about the objects. More specifically, there was a general complaint that, even with respect to those objects about whom the sorter felt well informed (from newspaper reports, biographies, etc.), the kind of information that becomes available concerning political personages was hardly appropriate for Q sorting a set of personological items. They protested, and quite rightly, that it was difficult to determine, for example, whether a public figure "lacks insight into his own motives and behavior." The general feeling was that the Q-sort items were inadequate for the expression of their real impressions about the objects. It might well be argued that our items were inappropriate for describing our objects, with consequent effects upon the relevance of our findings. In defense, we can say that all personality description is highly inferential. While our sorters were required to make their inferences from a very different (and perhaps less valid) class of data from that usually employed in Q-sorting situations, it seems reasonable to regard their judgment task as essentially comparable with the more conventional one. However, we can point to evidence which tends to vitiate the sorters' objections. Using the Cronbach and Gleser D statistic (3), we computed quantities describing the ten distances among the five objects from the composite Q sorts of the first run. Ranking these values, we obtained a picture, from Q-sort data only, of how our sorters saw each object in relation to every other object. We then asked our sorters to provide us with a more direct measure of this perceived distance between the objects. Using a set of ten cards, each designating one of the ten interobject comparisons, the sorters made an independent phenomenological ranking of "how far apart in your mind, how different from each other, are these five persons?" Although the sorters reported a variety of interpretations of the "difference" dimension (e.g., political orientation, personality type), they showed a significant agreement amongst themselves in rankings. Using Kendall's coefficient of concordance (6), we find a W of .43 ($p < .01$).[3] The rank-difference correlations between the Q-sort distance rankings (two, one each from forced and unforced data) and this phenomenological ranking are .75 and .71, respectively. Both of these values indicate significant relationships ($p < .05$). We can conclude from this that the Q sort, by this test, can adequately reproduce the phenomenology of the sorter. The Q sort does seem to be able to say what the sorter wants to say despite the sorter's doubts that his true impressions are "coming through."

Reports in the literature question the validity of self-evaluations of judgments when such judgments are not based on intensive, first-hand observation. In a study of behavior prediction from clinical materials, Polansky (7) found that confidence and felt adequacy of information correlated .008 and .088, respectively, with accuracy of judgment. Forer and Tolman (5), investigating prediction of responses on a sentence completion test, have demonstrated a significant inverse relationship between confidence in a judgment and agreement of the judgment with group consensus.

These findings led us to suspect that even the rankings by the sorters of the amount of

[3] As a secular but possibly interesting aside, it may be noted that our seven Democratic and four Republican sorters, in April of 1954, provided (from most to least different) the following composite phenomenological ranking: AS-JM, RT-JM, HT-JM, DE-JM, RT-HT, DE-HT, AS-RT, RT-DE, AS-DE, AS-HT.

information they had about the objects and of the confidence they felt in each sort might bear little relation to the reliability of their sorts. These rankings were compared with both forced and unforced reliabilities and no relationship whatsoever was found to exist, indicating again that self-evaluation is irrelevant to reliability of judgment.

The general import of the above discussion is that, while the Q sort is readily understood and performed, it tends to arouse considerable insecurity and, under the forced constrictions, resistance in the sorter. At face value, these reactions would seem seriously to discredit the information provided by these sorts. However, as we have shown, our sorters were generally incorrect in their evaluations of their own judgments. Two inferences may be drawn from these observations. First, there is the eminently practical suggestion that the sorter is his own worst critic and therefore the experimenter using the Q sort as an instrument should not be unduly alarmed by disparaging self-evaluations of the method by the sorters. The second implication, a much broader one, is that, insofar as the sorter is a poor judge of his own responses, the dictum—that the "comfortableness" and "naturalness" of the judging task are critical desiderata for acquiring valid data—should perhaps be re-examined.

SUMMARY

The present study investigates the effect of the frequency distribution of item placements upon the nature, amount, and reliability of the information contained in Q-sort personality descriptions. The relative merits of forced and unforced distributions are discussed and recent evidence is cited supporting the conclusion that, in most Q-sort applications, a prescribed forced distribution should be used. The appropriateness of the usual quasi-normal forced distribution is then questioned. The specific problem posed is to determine which form the forced Q-sort distribution should take.

Using a 76-item Q-sort set, eleven sorters made four sorts each describing the personalities of each of five political figures. In sorting each person, the sorter was first permitted to distribute the items among nine response categories as he wished. He was then required to rearrange the items to conform to a quasi-normal distribution. These two steps were repeated at a later time, thus obtaining four sorts of each of the five persons from each sorter. The reliabilities of both unforced and forced sorts were determined and the number of discriminations present in each unforced sort was computed.

The heterogeneity of the distributions employed for the unforced sorts indicate that no particular form of distribution was consistently preferred by the sorters. In view of this fact, the obtaining of maximum reliable information was adopted as the criterion for the optimum forced distribution. It was found that as more discriminations were made in the Q-sort situation, the test-retest reliability of the sort tended to increase. The data suggest that the maximum possible number of discriminations can be made among a set of items with no appreciable loss of reliablity. The recommendation is therefore offered that a rectangular distribution be employed for forced Q sorts.

Phenomenological reactions of the sorters to the sorting task are reported. Evidence is presented which not only indicates that these sorter reactions do not accurately reflect sorter performance, but also that the Q sort is an adequately sensitive instrument for the expression of personality descriptions.

REFERENCES

1. BLOCK, J. A comparison of the forced and unforced Q-sorting procedures. *Educ. psychol. Measmt*, 1956, in press.
2. CRONBACH, L. J. Correlations between persons as a research tool. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research*, New York: Ronald Press, 1953. Pp. 376–388.
3. CRONBACH, L. J., & GLESER, GOLDINE C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456–473.
4. FERGUSON, G. A. On the theory of test discrimination. *Psychometrika*, 1949, **14**, 61–68.
5. FORER, B. R., & TOLMAN, RUTH S. Some characteristics of clinical judgment. *J. consult. Psychol.*, 1952, **16**, 347–352.
6. KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.
7. POLANSKY, N. A. How shall a life history be written? *Charact. & Pers.*, 1941, **9**, 188–207.
8. SEEMAN, J., & RASKIN, N. L. Research perspectives in client-centered therapy. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research*. New York: Ronald, 1953. Pp. 205–234.
9. STEPHENSON, W. *The study of behavior*. Chicago: Univer. of Chicago Press, 1953.

Received February 21, 1955.

THE ROLE OF SMALL GROUPS IN MEDIATING THE EFFECTS OF PROPAGANDA

MAY BRODBECK[1]

University of Minnesota

An individual exposed to persuasive communication from a supposedly reputable source may, broadly, react in any of three ways: (*a*) If he was previously disposed in favor of the view presented, then his confidence in his opinion is either strengthened or it remains unchanged. (*b*) If he was previously opposed to the view presented, then his confidence in his opinion is either weakened or remains unchanged. (*c*) If he was previously opposed, then he may change to the other side. Each of these alternatives presumes that the source of the communication is respected by the hearer; if this is not so, there may well be quite contrary effects in each case.

Suppose, then, that a group of individuals, some opposing and some supporting a particular view, are exposed to a communication that strongly favors only one side of the issue. There will be individuals whose views are reinforced by the communication and, if it is effective, others whose confidence in their opinion is in some degree shaken. Let us consider for the moment only these two cases. Suppose further that all these individuals are then encouraged to engage in a free discussion in which whoever wishes may speak on the issue, express agreement or disagreement with the communication, ask questions of others, etc. It is reasonable to expect that those who have just heard forcible arguments against their own views will be unpleasantly disturbed and will make use of the discussion to allay their uneasiness. If, in the discussion, they are more attentive to participants on their own side than they are to those on the other side they may achieve this result. The following hypothesis is therefore suggested: Given a choice, individuals whose confidence in their belief has been shaken by a communication subsequently prefer to listen or speak to those who agree with them, while those whose opinion has been reinforced subsequently show a greater tendency to communicate with those on the other side. Since greater attentiveness to the views of those who agree with them has the effect of counteracting the propaganda, it follows as a corollary that given opportunity to participate in a free discussion group, those shaken individuals who show this greater tendency to seek support for their beliefs tend to revert to the degree of confidence expressed before exposure to the communication. In this situation they may be expected to tend to listen to arguments presented by their own side and ignore arguments of the opponents. It should be noted that the absolute level of confidence is not relevant to the hypothesis, only a shift in strength of belief.

There remain two more cases to be considered among the possible reactions of persons who hear a speech opposed to their own views. Some individuals will not express a decrease in confidence after exposure. For these people either the experimental manipulation has not been successful or they are reluctant to admit that it has been. Their behavior may be expected to fall, on the average, between that of those who do express a decrease in confidence and the behavior of those whose views have been reinforced by the propaganda. They should exhibit a smaller tendency to choose people who agree with them for face-to-face communication in the discussion situation than do the shaken individuals, but a larger tendency to do this than do those whose views were supported by the communication. The other remaining case is that of the people who change to the other side. They will wish to communicate with those who are *now* on the same side as they are, that is, the side they newly adopted as a result of the communication. In other words, the desire to communicate with individuals on the same side is an increasing function of the degree to which an individual's opinion is shaken by the propaganda, up to the point where conversion takes place. The converts, however, will also be

[1] This study was done in conjunction with the Laboratory for Research in Social Relations. The author is indebted to Professor Leon Festinger for considerable assistance in the conception and design of this experiment and to Alvin Boderman and Danuta Ehrlich for able technical assistance in its execution.

insecure in their opinions and, unlike those who initially agreed with the persuasive communication, they will seek to communicate with people on what is now their own side.

In sum, the hypothesis states that if the strength of an individual's conviction is merely weakened by propaganda, then the effect of the propaganda tends to be nullified by later participation in free-discussion situations since he seeks out those who oppose it. To be effective, the communication must be persuasive enough to cause complete conversion, for in this case he tends to seek discussion with others who support the propaganda and reinforce its effect. The following experiment was designed to test the hypothesis and corollary.

PROCEDURE

University of Minnesota students were recruited from classrooms as volunteers for a study of attitudes. Both male and female students were permitted to volunteer, although no use was made of female volunteers until late in the experiment. The pretests revealed that undergraduate women were insufficiently interested in the topic selected for experimental manipulation. Five groups of graduate women were used, however, and since their behavior appeared to be no different from that of the men students, all the data have been combined. From 12 to 15 *Ss* were called for each group. Care was taken when selecting *Ss* and at the experimental session to ensure that the members of the group were relative strangers to each other. This precaution was necessary to minimize the effect of extraneous considerations on the *Ss*' choices. The *Ss* were not encouraged to speak to each other before the experiment began. It was not, however, always possible to prevent this. No communication was permitted among the *Ss* once the experiment began.

In outline, the sequence of the experiment was as follows:

1. The *Ss* filled out a questionnaire stating their opinion and level of confidence on an issue.
2. They were then exposed to propaganda supporting one side of the issue.
3. After exposure, they again filled out the opinion questionnaire.
4. They were then asked to list on a choice questionnaire their first, second, and third choices of people present, whose views were made public, whom they would like to listen to or speak with on the issue.
5. The entire group engaged in open, free discussion of the issue.
6. After the discussion, the opinion questionnaire was again filled out. The details follow.

Upon entering the experimental room *Ss* were given numbered identification slips. When sufficient *Ss* had appeared, preferably at least 12 (but some groups began with as few as eight or nine), they were told that the Laboratory for Research in Social Relations was participating in a nationwide study for a (fictitious) National Council for Adult Education and that they would later discuss the issue of wire tapping but that we should first like to have their opinions on this topic. The first Opinion Questionnaire was then distributed and *Ss* were asked to write their numbers at the top of it. The questionnaire was as follows:

> "A law now exists prohibiting the tapping of telephone wires and the use of information gained by wiretapping as evidence in court. This law is not enforced and much wiretapping by private individuals and the government does go on. There is now a proposal before Congress to permit the government to engage in wiretapping. Some people believe that the government should have this right. Others oppose it on the grounds that if wiretapping is made legal for the government, then private individuals and non-governmental agencies will also engage in wiretapping. Will you please state your opinion of this issue by checking the appropriate blanks below.
>
> 1. Wiretapping is an invasion of privacy and should continue to be illegal for everyone, including the government. Please check one: Yes, No.
> 2. Please check one of the following: (a) I am absolutely certain of the opinion I expressed above. (b) I am quite certain of the opinion I expressed above. (c) I think that the opinion I expressed is right, but I am not certain about it. (d) I'm a little doubtful, but I think my opinion is right. (e) I'm quite doubtful, but am not just guessing. (f) I'm completely uncertain about whether my opinion is right or wrong."

While collecting the questionnaires, an assistant counted the number of *Ss* on each side of the question. Meanwhile, the experimenter told *Ss* that in order to stimulate discussion, it would be a good idea if they heard some arguments that could be given one way or the other on the issue. For this purpose, we were fortunate in having on tape a (fictitious) speech given by the (equally fictitious) President of the National Council for Adult Education. They were told that the speech would take about eight or nine minutes, as it did, and asked to listen carefully. Two speeches were always ready on the same tape, each giving a strong argument for one side of the issue. The speech presented always opposed the opinion of the majority of the students in the group. If there happened to be an even division, then the speech for which the tape was set was used. (It had been previously ascertained from questionnaires containing ten questions on various issues, distributed in large classes from which *Ss* were not recruited, that student opinion was about evenly divided on the issue. The level of confidence on this question was moderate and there had not been much discussion by students about the issue. Both these things were desirable in order to maximize the chance of success of the experimental manipulation.)

While *Ss* were listening to the speech, a list was made of their opinions and levels of confidence, for easy comparison with the second questionnaire. After the speech was over, *Ss* were asked to fill out the second questionnaire, which for the first 17 groups was identical with the first one. (Close to the end of the experiment, after 17 groups had been run, in order to increase the number of *Ss* who admitted to a decrease in confidence as a

consequence of exposure to the speech, it was considered advisable to change Questionnaire 1 by adding three irrelevant questions in order to minimize the effect of memory and reluctance to admit change. In this case *Ss* were told after the first questionnaire was collected that while the other information was valuable, wire tapping was the issue we would be concerned with at this time.)

After the second questionnaire was collected, the experimenter told *Ss* that an identical experiment was going on in another part of the building and that some of them would be asked to join this other group after the assistants had "randomized" their numbers. The experimenter then spent some time telling them about different kinds of discussion, such as group discussion, discussion in pairs, leader-led discussion, etc. She also mentioned a type of "discussion" using a one-way intercommunication system, with which an individual could listen or speak to another person but not both. The main purpose of this dissertation was to hold the students' attention while the second questionnaires were compared with the results of the first by an assistant, in order to form a group of eight *Ss* composed of four controls and four experimental *Ss*. After the group had been formed by an assistant, the experimenter's dissertation on varieties of discussion was interrupted and the assistant called out the excess *Ss* by number, presumably picked at random, to join the "other group." An assistant accompanied them outside, explained the study, answered questions, and dismissed them.

In summary, the eight *Ss* remaining in each final group consisted of the following:

C: *Ss* whose opinion on the second questionnaire agreed with the opinion of the influence communication. (In all cases except five these *Ss* initial opinions were also on that side. These five, who had switched sides after hearing the propaganda, were included only where necessary to obtain four C *Ss*. There was one such convert in each of five groups.)

E: *Ss* whose initial attitude disagreed with the opinion of the communication and who indicated a decrease in confidence after exposure. There were never as many as four such *Ss* and the balance was filled out with E*s.

E*: *Ss* whose initial attitude disagreed with the opinion of the communication but who did not express a decrease in confidence after exposure.

There were four "agreers" and a total of four "disagreers" (E and E* counted together) in each group.

Meanwhile, the experimenter explained to each group that they might be chosen to use the one-way telephone type of discussion. Should this occur, we would like them to have the opportunity of designating to whom they should like to listen, if they were on the listening end of the apparatus, or speak, if they were on the speaking end. We would publicly identify their opinions so that they could make their choices intelligently. Accordingly, a large card was placed in front of each *S*, on which his number and his position, "Yes" or "No," was written. (His level of confidence was *not* revealed.) Each *S* could see everyone else's card and it was clear that the group was evenly divided. The choice questionnaire was then distributed and *Ss* were asked to indicate their first, second, and third choices for both listening and speaking.

After the choice questionnaire was collected, the experimenter pretended to sort them. *Ss* were then told that because of so much overlapping in choices, it would be too difficult to arrange the telephone situation. Instead, they would engage in group discussion among themselves. They were further told that we would observe the discussion, but would not participate in it. No attempt was made by the experimenter to get the discussion started. They were permitted to discuss the issue for twelve minutes, after which the experimenter brought it to a close and Opinion Questionnaire 3, identical with the second one, was distributed. When these were collected, the experiment was completely described, all the deceptions were revealed and the reasons for them explained, questions were answered, and *Ss* dismissed.

Summary of Measures Obtained

1. Initial attitude toward wire tapping, and degree of confidence.
2. Attitude toward wire tapping, and degree of confidence, after hearing speech.
3. Preferences for listening or speaking partners. The choice questionnaire obtained the first, second, and third choices for both listening and speaking.
4. Final attitude toward wire tapping, and degree of confidence, after free discussion.

RESULTS

Success of the Experimental Manipulation

Data were obtained from 22 groups. Two groups were discarded because of errors, leaving 20 groups or a total of 160 *Ss*, 80 experimentals, and 80 controls. Of the former, 28 indicated a decrease in confidence. These are designated E. The remaining experimental *Ss* are designated E*. There was no significant difference in the number of E's created by the two speeches. One E and one E* were discarded because of errors of procedure, and five groups were used, each containing one *S* who had changed to the other side. The final data therefore are based on 27 E, 51 E* and 75 C.

Subjects' Choices for Communication

Previous studies (1, 2) have shown that in discussion situations people have a greater tendency to communicate with those who disagree with them than with those on their own side. In the present experimental situation, the chance probability of each *S* choosing someone on the opposite side in both first and second place is only 2/7 or 29 per cent. As Table 1 shows, the actual tendency is considerably greater in all cases, a finding in agreement with the results of the earlier studies.

But by hypothesis, the experimental manipulation, exposure to a persuasive communication with consequent decrease in confidence, introduces another factor in opposition to the tendency expected on the basis of the previous studies. The experimental *S*s should show less tendency to choose those who disagree with them than the controls. As contrasted with the random probability, which is the same for both groups, the actual percentages are 65 per cent for the C's and only 44 per cent for the E's. All data are given for the Listening choices. While in the same direction, the data for the Speaking choices are statistically less significant. This difference was expected on theoretical grounds, since the E's are presumably not interested in talking, so much as in listening to arguments for *their* side. Forced as they were to a choice of whom to speak to, they may nevertheless be expected to show a tendency to choose people on the other side and talk about the position which, after all, is the one they still support.

The basic data for testing the main hypothesis are, then, the number of choices of listening partners who agree with them made by the E's and by the C's. The behavior of the E*'s, those *S*s who disagreed with the speech but who indicated no loss of confidence, can be expected to lie between that of the E's and the C's, as was in fact the case. Table 1 shows the data for choices by the *S*s of listening partners who agree with them. The difference between the frequency with which E's and C's assigned their first choices to their own side is significant at the 7 per cent level of confidence, when tested by chi square. The same test applied to frequencies based on first or second choices or both shows the difference to be significant at the 6 per cent level. Two-tailed tests were employed in both instances. (Because of the larger *N* involved, data are given henceforth for those who made either their first or second choice, or both, on the same side, rather than those who made only their first choices in this way.)

TABLE 1

CHOICES OF LISTENING PARTNERS

Choices	Types of Subjects			
	E	E*	C	
First choices				
Own side	9	11	11	E vs. C, $\chi^2 = 3.281$
Other side	18	40	64	$p = .07$
First or second choices				
Own side	15	21	26	E vs. C, $\chi^2 = 3.608$
Other side	12	30	49	$p = .06$

TABLE 2

EFFECT OF DISCUSSION

Effect	Types of Subjects		
	E Chose Own Side	E Chose Other Side	C Changed by Speech
Return to initial confidence level			3
Over-recovery	5	1	
Complete recovery	6	2	
Do not return			20
Partial recovery	0	2	
No recovery	4	7	

E chose own side vs. E chose other side, $p = .03$ by exact test.

E chose own side vs. C, $\chi^2 = 11.7$, $p < .01$.

The Recovery Phenomenon: Effects of the Discussion

The corollary predicts that after participation in free group discussion, those *S*s whose confidence has been lowered by the speech tend to revert to their original level of confidence. The possibility exists that this represents a tendency unrelated to our experimental manipulation. If such a general "snap-back" phenomenon accounts for the results, however, there should be no difference in this respect between the "E chose own side" and the "E chose other side" groups, nor between the former and the controls who increased their confidence after hearing the speech. The data are to the contrary. Particularly relevant to the hypothesis is the difference within the E group. Those E's who do not choose those who agree with them are also those who are relatively unaffected by the discussion. Not oriented toward getting support, as is indicated by their choices, they also do not "listen selectively" during the discussion. Some other factor, possibly indifference to the issue, is apparently operative.

This experiment cannot support any assertions about what would happen in the absence of discussion. A test of the corollary to the hypothesis, however, does not depend upon

the introduction of a new experimental condition. All *S*s were undifferentially exposed to the discussion. But the corollary asserts that those shaken experimental *S*s who, on the choice questionnaire, reacted differentially with respect to seeking support also tend to react differentially to the discussion. The greater tendency to choose people on their own side is interpreted as seeking support which, in turn, leads us to expect selective listening during the discussion. The hypothesis asserts that the extent of reversion to initial level of confidence should be different for the E's choosing those who agree with them than for the E's who do not do so and, for that matter, for those C's who increased confidence after exposure. The data obtained from the choice questionnaire permits the reversion predicted by the corollary to be explained in terms of seeking support. The discussion situation simply offers one immediate way these *S*s can obtain that support.

Table 2 gives the comparison between the E's whose first or second choices or both were made to people who agree with them, the E's who did not make their choices in this way, and also the Controls whose confidence level was changed by the speech. Seventy-three per cent of the E's who chose individuals on their own side revert to their original level of confidence as compared with only 25 per cent of the E's who did not choose their own side and 13 per cent of the C's. The statistical comparison is made between those who show either no change in confidence from the second to the third questionnaires or only partial recovery, on the one hand, and those who show complete or over-recovery after the discussion, on the other. By exact two-tailed probability test, the difference between the E's who chose agreers and the E's who did not do so is significant at the .03 level. The difference between the former and the C's is significant at better than the .01 level of confidence when tested by chi square. Clearly, the E's who chose people on their own side are strongly affected by the discussion while all other subjects are relatively indifferent to it. Both the differential influence of the discussion on the E's who choose those who agree with them and the direction of this influence are as predicted by the hypothesis.

Subjects Completely Converted

Something should be said about those *S*s who change to the other side after hearing the persuasive communication. It was predicted that the behavior of such individuals would be similar to that of the E's in that they would tend to choose people on the same side as themselves, but in their case the "same" side would be that of the side they had newly adopted, rather than the side they initially professed. Of the five *S*s used who shifted sides after exposure, four made their first or second choices on the same side, that is the side to which they changed. Although the number of cases is too small to warrant any firm conclusion, the converts appear to behave in the manner predicted.

SUMMARY

This experiment was designed to test the related hypotheses that (*a*) individuals whose confidence in a belief has been shaken by exposure to propaganda opposing their beliefs prefer to hear arguments from their own side in order to bolster their confidence and (*b*) that given opportunity to participate in a free discussion group after the exposure, these individuals tend to listen preferentially to persons who agree with them and to ignore the arguments of their opponents, with the consequence that their confidence in their opinion tends to return toward its initial level. In other words, the hypotheses state that participation in free group discussion tends to counteract the influence of the communication. The data support the hypotheses.

REFERENCES

1. FESTINGER, L., & THIBAUT, J. Interpersonal communication in small groups. *J. abnorm. soc. Psychol.*, 1951, **46**, 92–99.
2. GERARD, H. B. The effect of different dimensions of disagreement on the communication process in small groups. *Hum. Relat.*, 1953, **6**, 249–271.

Received February 23, 1955.

SELF RATINGS OF FEAR IN A FEAR-INVOKING SITUATION

RICHARD D. WALK[1]

Cornell University[2]

A SELF-RATING scale of fear may offer research possibilities as an approach to the measurement of how fearful a situation is to the individual. A first requirement of such a scale is that the self ratings must be related to some correlates of fear. One aspect of the present study is the relation between fear on the rating scale and performance in a fearful situation. Another topic is the relation between self ratings and other verbal reports of fear. From performance, an inference of the validity of the rating scale may be made; from other questions on fear, reliability (in the sense of generality) may be estimated. If both reliability and validity can be inferred, another possible step is to study the attitudinal factors that predispose some subjects to give high and some to give low fear ratings.

Predisposition to admit fear when exposed to a fearful situation is one aspect of susceptibility to stress. The general topic of psychological stress has been carefully reviewed elsewhere (8). Susceptibility to stress is particularly considered in Air Force (9) and OSS (11) studies during World War II. Neither program developed adequate measures of susceptibility to stress that were related to later performance. The Air Force developed such tests as the Multiple Control Stress Test, the Observational Stress Test, and the Steadiness Under Pressure Test. The report concludes:

> The available data do not support the hypothesis that additional validity for the prediction of success in elementary pilot training accrues to a test situation when verbal threats and other distractions, including presumably fear-producing stimuli, are administered. . . . It must be recognized that extreme forms of emotion-producing stimuli were either not attempted or were not permitted during the course of the exploration of this type of test in the Aviation Psychology Program (9, p. 849).

The OSS program was self-validating in the sense that performance in the experimental sessions was used as a basis for accepting or rejecting candidates. It is admitted, however, that the program was not markedly successful in predicting subsequent performance in the field. Recently, Williams (16) claimed that certain Rorschach variables were associated with susceptibility to stress. Other experiments have failed to confirm his results (3, 4, 13).

For the study of fear, the military training situation offers more realistic stress than it is often possible to produce in the laboratory. Airborne trainees are required to learn to jump from a 34-foot mock tower. The trainee jumps from an exit door 34 feet high and falls nearly 8 feet before his fall is snubbed by straps fastened on the rear of his parachute harness. The straps extend to a pulley arrangement on a steel cable and, after his fall, the trainee rides the cable approximately 75 yards before he reaches a mound and is unhooked. The purpose of the mock-tower training is to teach the trainee proper exit form from an airplane in flight. A member of the cadre grades trainees on their jump performance, and those who cannot learn proper jump form fail the Airborne training course. The training procedure is generally held to be frightening (14, pp. 212–220), and many trainees react to it with such overt fear signs as extreme pallor or marked trembling. The OSS assessment staff described a similar training procedure in India during World War II as "an exceedingly rigorous test which evoked marked emotional responses in all candidates" (11, p. 370). Similarly, natives in the New Hebrides jump from high towers over land as a courage test to prove their manhood (7). Their fall is snubbed by vines.

Two studies using a rating scale are reported. Study I was concerned entirely with the relation between fear ratings and performance at the 34-foot mock tower. Trainees were required to rate the amount of fear they felt just *prior* to a jump from the mock tower.

[1] The author is indebted to Drs. John L. Finan and Richard L. Solomon for their assistance in planning the research, to Ted Stephens for collecting the data for Study I, and to Janet Heilmann for her aid in analysis.

[2] This study was performed while the author was employed by the Human Resources Research Office, The George Washington University, operating under contract to the Department of the Army. Opinions and conclusions are those of the writer and do not necessarily represent views of the University or the Department of the Army.

Study II, on the other hand, dealt with performance at the mock tower and also with attitudes. The men in Study II were given an attitude questionnaire prior to the beginning of Airborne training. They were then asked to rate, at the conclusion of mock-tower training, how afraid they felt as they made their first five jumps. Other questions asked at the conclusion of mock-tower training helped check on the reliability of the self rating of fear.

PROCEDURE

The Instrument

The trainee was asked to place a mark across the sketch of a thermometer-like figure to indicate how much fear he felt. The method of marking was explained so that high fear was toward the top of the thermometer, and low fear was toward the bottom. A permissive attitude was expressed toward the admission of fear during the orientation. The experimenter said, "We expect that you will be afraid on some of your jumps." To measure fear, the thermometer was divided into 10 equal parts. The distance from the base of the thermometer to the pencil mark made by the trainee thus yielded a numerical index which ranged from 1 through 10 and represented the "amount" of fear felt by the trainee.

Mock-Tower Rating Procedure

Every time a trainee jumped from the mock tower, a cadre member rated his jump on its approximation to satisfactory (prescribed) jump form. Errors were recorded on a rating sheet, and the trainee was told the errors he had made to aid him in correcting his faults. Trainees made from 10–25 mock-tower jumps in the first week of training. To study the relation between fear and performance, the mock-tower jump on which a trainee made his first satisfactory performance was recorded, as was the number of errors a trainee made on each jump. Errors ranged from zero errors (a satisfactory jump) to five, the most that were recorded by the raters. Trainees were not graduated from the first week of training until they made three satisfactory jumps from the mock tower. The records of a trainee's performance at the mock tower were available for this investigation from records kept by the Airborne Department.

Subjects and Procedure for the Two Studies

Study I. Only those trainees who completed a rating scale on their first seven mock-tower jumps were retained in the sample. One squad of men (15–20 trainees) was studied every week for seven weeks, and, of the 107 men on whom complete records were available, 76 later passed the Airborne course with 31 failing it. The thermometer-like rating scale was handed to a man just prior to the time he made a mock-tower jump. The records available for Study I were, thus, the self ratings of fear for seven jumps and the records of performance (error ratings for each jump and ultimate pass-fail status).

Study II. The sample for this study was two classes of Airborne trainees, and only trainees who took the questionnaire before the beginning of Airborne training and again at the end of mock-tower training were used in the analysis. A total of 313 trainees had complete records.

The first questionnaire consisted of background items, a sentence completion test, and direct multiple-choice questions. Included in the direct questions was a number of items on psychosomatic reactions. In the second questionnaire trainees were asked to rate, in retrospect, the amount of fear they felt on their first five mock-tower jumps. The same thermometer sketch as the one in Study I was used, and a total of five separate sketches were supplied to represent the first five mock-tower jumps the trainee made from the mock tower. After the thermometer sketches were filled out, the *S*s completed a series of direct questions about fear reactions.

Airborne performance records (mock-tower performance and pass-fail status) were available for these *S*s, and Army IQ scores (the AFQT scores) were obtained from army records for most of them.

RESULTS

Study I

Since introspective fear scores were markedly skewed toward the bottom of the scale, particularly after the first two mock-tower jumps, medians are used to present the data.

The self ratings of fear begin fairly high on the first mock-tower jump and rapidly fall with successive mock-tower jumps. This is shown in Fig. 1. The decline is highly significant ($\chi^2 = 32.82$, $p = .001$). Figure 1 also shows that those who later failed the Airborne course appear to admit more fear than those who passed it, but none of the differences is statistically significant (.05 level).

The men who later passed the Airborne course were divided into quartiles on the basis of when they received their first satisfactory (errorless) jump rating from the mock tower. The distribution is shown in Table 1. Median fear scores for these groups of satisfactory jumpers are presented in Fig. 2. The group that achieved a first satisfactory jump very early in training gave lower median fear estimates than those whose first satisfactory jump was very late in training. On four of the seven mock-tower jumps (Jumps 2, 4, 6, and 7), the differences in the self rating of fear between these two groups are significant (.05 level). The group of early and late satisfactory jumpers are not statistically different from each other or from the two extreme groups.

The correlations between the number of er-

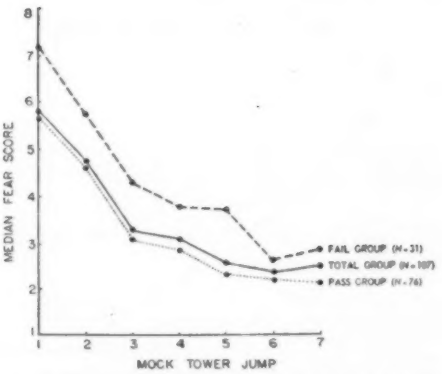

FIG. 1. MEDIAN SELF RATINGS OF FEAR FOR THE TOTAL GROUP AND FOR THOSE WHO LATER PASSED AND THOSE WHO LATER FAILED THE AIRBORNE COURSE (STUDY I)

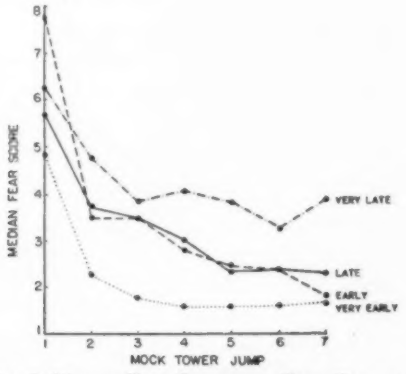

FIG. 2. MEDIAN FEAR SCORES FOR FOUR GROUPS OF SATISFACTORY JUMPERS, ALL OF WHOM PASSED THE AIRBORNE COURSE (STUDY I)

TABLE 1

TRAINEES OF STUDY I DIVIDED INTO QUARTILES ON THE BASIS OF MOCK-TOWER PERFORMANCE

Quartile	First Mock-Tower Jump Judged Satisfactory	*N*
Very early satisfactory group	Jumps 1–5	17
Early satisfactory group	Jumps 6–8	21
Late satisfactory group	Jumps 9–11	21
Very late satisfactory group	Jumps 12–21	17
Total		76

TABLE 2

CORRELATION BETWEEN FEAR RATINGS AND ERRORS ON MOCK TOWER FOR STUDY I ($N = 107$)

Mock-Tower Jump	r*
1	.12
2	.28
3	.42
4	.44
5	.44
6	.38
7	.45

* An *r* of .20 is required for significance at the 5% level and an *r* of .26 at the 1% level.

rors made by trainees on each jump and the trainees' fear ratings made just prior to that jump are shown in Table 2. With the exception of the correlation for the first mock-tower jump, where the fear rating was made prior to the time a trainee had actually jumped from the tower, all *r*'s are significant at the 1% level.

To conclude, then, this study shows a fairly close inverse relationship between a self rating of fear and adequacy of performance in a fearful situation. Those who perform well rate themselves as low in fear; those who perform poorly rate themselves high in fear.

Study II

Retrospective fear ratings and performance. Since the retrospective fear ratings are spread more over the scale than those made immediately prior to mock-tower jumping, means are presented rather than medians. Twenty-two of the 313 trainees who took both questionnaires were also in Study I. Their ratings are similar to those who first used the self-rating scale after the completion of mock-tower jumping, but they were not considered in the analysis of retrospective fear ratings. They were, however, retained in the analysis of questionnaire items.

By the time in training when these trainees were given the rating scale, after the completion of mock-tower training, many of the trainees who failed the course had already been removed from the training class. Of the 291 trainees who took both questionnaires, 243 later passed the Airborne course, while 48 failed it. The self ratings of fear for those who later passed or failed are shown in Fig. 3. Those who later failed the Airborne course admitted significantly more fear than those who passed it (.05 level).

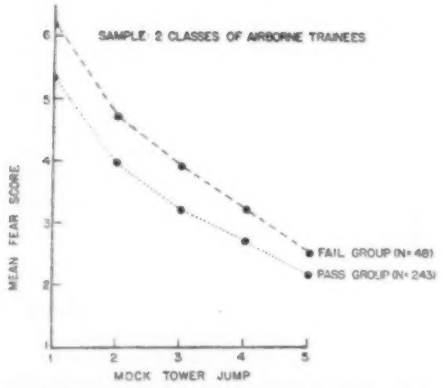

FIG. 3. MEAN RETROSPECTIVE SELF RATINGS OF FEAR AT THE END OF MOCK-TOWER TRAINING FOR THOSE WHO LATER PASSED AND THOSE WHO LATER FAILED THE AIRBORNE COURSE (STUDY II)

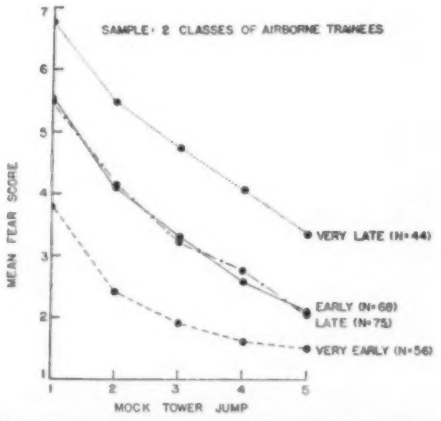

FIG. 4. MEAN RETROSPECTIVE SELF RATINGS OF FEAR AT THE END OF MOCK-TOWER TRAINING FOR FOUR GROUPS OF SATISFACTORY JUMPERS, ALL OF WHOM PASSED THE AIRBORNE COURSE (STUDY II)

TABLE 3

TRAINEES OF STUDY II DIVIDED INTO QUARTILES ON THE BASIS OF MOCK-TOWER PERFORMANCE

Quartile	First Mock-Tower Jump Judged Satisfactory	*N*
Very early satisfactory group	Jumps 1–5	56
Early satisfactory group	Jumps 6–8	68
Late satisfactory group	Jumps 9–10	75
Very late satisfactory group	Jumps 13–21	44
Total		243

As in the previous study, the trainees who later passed the Airborne course were grouped according to their first satisfactory jump rating at the mock tower (see Table 3). The relationship between performance and fear rating is shown in Fig. 4. Trainees who performed the best in training (very early satisfactory group) gave lower fear ratings than those who also passed but performed the poorest (very late satisfactory group). The differences between the two groups are statistically significant (.001 level) on every mock-tower jump. As in Study I, the early and the late satisfactory group are indistinguishable from each other, but the two groups (combined) are significantly different (.02 level) from either the very early or the very late group on every jump.

Self ratings of fear and other measures of fear. One of the first questions that can be asked about the mark on the self-rating scale is whether this mark has any relation to other measures of fear. It can be maintained that people differ in the way they perceive a self-rating scale and that, subjectively, a mark at the high end of the scale for one person may be phenomenologically equal to a mark on the low end for another. The second questionnaire, therefore, included questions on physiological reactions for the previous few days and also asked whether the trainees had been afraid. This latter question was phrased as follows:

(Question 7) DURING THE PAST FEW DAYS would you say you were really scared? (Check one)

___Many times
___A few times
___Once
___Never

To compare the answers on the direct question with the marks on the self-rating scale, the fear ratings on the first mock-tower jump were used. The 10-point rating scale was divided into fifths, Intervals 1–2 representing the low end of the scale and Intervals 9–10 the extreme high end. Fig. 5 shows that of the trainees who marked the lowest fifth of the scale (Intervals 1–2), 90 per cent answered, on the direct question, that they had "never" been scared in the previous few days. Each succeeding fifth of the scale contains a greater proportion answering

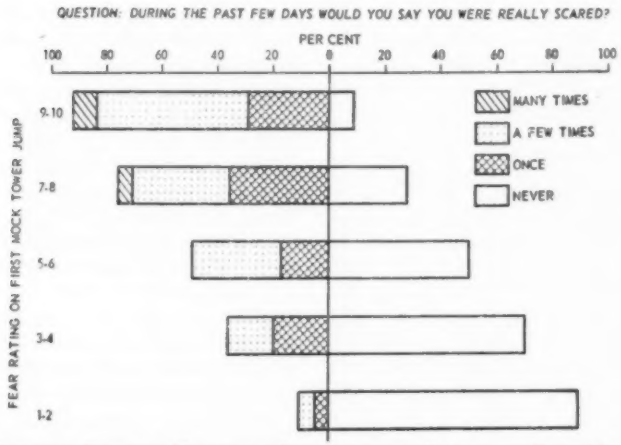

FIG. 5. THE RELATION BETWEEN NUMERICAL SELF RATINGS OF FEAR ON THE FIRST MOCK-TOWER JUMP AND A DIRECT QUESTION ON BEING SCARED

that they had been "scared," and 90 per cent of those who marked the high fifth of the scale, Intervals 9–10, replied that they had been scared once or more times.

Although the direct question and the rating scale appear to be related, how do we know that the fear was induced by the training and was not some general personality tendency to admit fear? Prior to training, the trainees were asked: "How many times in your life would you say you have been really scared?" This item permits direct comparison of responses to a question given before and after mock-tower training as it is related to fear on the rating scale. Those who answered that they had "high" fear on the rating scale for the first mock-tower jump (Intervals 6–10) said, prior to training, that they had been frightened more often in their lives than the group that admitted "low" fear on the rating scale (Intervals 1–5). These initial, pretraining differences between the two groups could have happened only once in five times by chance ($\chi^2 = 2.69$, $p = .20$). After a mock-tower training, the group that admitted to high fear on the rating scale for the first mock-tower jump said, on the direct question, that they had been "scared" more often in the previous few days than did the low-fear group. But now the differences between the two groups could have happened by chance less than once in 100,000 times ($\chi^2 = 45.73$, $p = .00001$). Apparently, then, the fear reaction was primarily induced by the training itself and did not reflect a general tendency to admit fear.

Further support for this contention comes from the questions on physiological reactions. Both before and after mock-tower training, the trainees were asked about the occurrence of the following physiological reactions: hands trembling, nervousness, heart beating hard, shortness of breath, hands sweating, upset stomach, and cold sweats. Before training, none of these questions showed any significant differences between those who later gave high and those who later gave low self ratings of fear on their first mock-tower jump. In Table 4, the percentage of reported physiological reactions for the previous few days of high- and low-fear groups is shown. "Nervousness" and "heart beating hard" were reported significantly more often by the high- than by the low-fear group. Ax (1) recently showed that a high heart rate was one of the physiological reactions that differentiated fear from anger. A violent pounding of the heart was the most frequent reaction to combat reported by Army (14, p. 201) and Air Force (5, p. 214) veterans of World War II.

Although one must agree that a given pencil mark on the self-rating scale may not mean the same thing for every trainee, this study shows that, despite this factor, those who rate themselves high in fear on the self-rating scale also admit more often to fear on a direct question and indicate the presence of more physio-

TABLE 4

PERCENTAGE OF PSYCHOSOMATIC REACTIONS TO MOCK-TOWER TRAINING BY HIGH- AND LOW-FEAR GROUPS

Psychosomatic Reaction	Low-Fear Rating (n = 155)	High-Fear Rating (n = 158)	χ^2
1. Hands trembled	12%	17%	1.88
2. Nervousness	31%	44%	6.16*
3. Heat beat hard	24%	35%	4.56*
4. Shortness of breath when not exercising	12%	9%	.77
5. Hands sweating so they felt damp and clammy	21%	26%	1.30
6. Upset stomach	22%	24%	.22
7. Cold sweats	8%	9%	.06

* $p < .05$.

logical reactions during mock-tower training than those who rate themselves low in fear.

Questionnaire items and fear ratings. The focus of interest in this section is the antecedent attitudes or other factors related to a high or low fear rating. Physiological reactions checked on the first questionnaire, as was mentioned previously, were not related to a high fear rating for mock-tower training, and neither was the Army measure of general intelligence, the AFQT score.

What was related to a later high self rating of fear? The clearest relationships seem to be in reaction to possible danger, worry about injury, and confidence in ability to do well in a dangerous situation. For example, trainees were asked: "If a small group of volunteers were needed to carry out a dangerous mission, do you think you would volunteer?" Significantly more (.02 level) of those who later rated themselves as low in fear during mock-tower jumping answered, "I would certainly volunteer." A sentence-completion item illustrates the same trend. The item was, "When they said it was dangerous, Bert . . ." Out of 313 trainees, over 60 per cent ($N = 198$) gave a response in the categories either of a positive response to danger (e.g., "went on anyway," "didn't mind") or of a negative response (e.g., "quit"). Significantly more (.05 level) of those who gave a negative response rated themselves after mock-tower training as high in fear.

A number of questions on confidence and worry about injury showed significant differences between high- and low-fear groups. The items were as follows:

"If you were sent into actual combat now, how do you think you would do?" (Low-fear rating: "all right," $p = .001$).

"Do you ever worry about whether you will be injured in combat?" (Low-fear rating: "never even think about it," $p = .05$).

"If you were to make a parachute jump tomorrow, how well do you think you would do?" (Low-fear rating: I think I would do fine," $p = .01$).

"How hard do you think it would be for you to jump out of an airplane with a parachute?" (Low-fear rating: "very easy," or "easy," $p = .01$).

"Do you ever worry about whether you will be injured in Airborne training?" (low-fear rating: "never even think about it," $p = .01$).

DISCUSSION

A contribution of this study is its demonstration of the reliability and validity of a simple self-rating scale of the emotional reaction of fear. Reliability of the scale was shown in its high degree of relationship to responses to a direct question on being afraid and from the congruence of high-fear ratings and physiological fear reactions. Validity was inferred from the high-fear ratings of those who performed poorly in mock-tower training.

Since it has a certain degree of reliability and validity, perhaps the self-rating scale may be regarded as an indirect measure of an intervening variable, fear. This notion suggests the organization of the present findings in the manner suggested by Brown and Farber (2). An antecedent condition for the admission of fear is lack of experience at the mock tower. Fear ratings are initially high and with increased experience (mock-tower jumping), the fear ratings rapidly fall. Another antecedent factor for the expression of fear is an attitudinal factor concerning confidence; those without such confidence later express high fear, whereas those with prior confidence express low fear when exposed to the stress of mock-tower training. The consequent side of the formulation can be inferred from the study of the performance of those with high or low fear ratings. Those who admit high fear perform more

poorly on the mock tower and are more likely to fail the Airborne course than those who rate themselves as low in fear. This antecedent-consequent formulation fits in with other data. Shaffer (12), for example, reported that the three most frequently cited factors leading to decreased fear in combat fliers concerned confidence in equipment, crew, and one's immediate superior. Many trainees who passed the Airborne course, one may hypothesize, learned self-confidence on such training aids as the mock tower.

The questionnaire items which were related to a high-fear rating have implications for studies of decrements in performance under stress. Several previous studies have related questionnaire-inferred anxiety to decrements in performance under stress (4, 6, 10, 15). These studies have shown little relationship between questionnaire-defined anxiety and a decrement in performance under experimental stress. The present study showed no relation between previous anxiety and admission of considerable fear. Questions specific to danger, however, such as worry about injury in Airborne training or confidence in ability to do well in combat or parachute jumping, did show significant relationships (a nonsignificant relation between confidence and improvement in performance under stress was reported in [4]). The prior attitudes, thus, were related to the type of stress itself and not to a more general "anxiety" reaction. An implication of this study is, then, that questionnaire items on reactions to stress might be related to performance under stress.

This interpretation, however, deserves a word of caution. A high-fear rating for mock-tower training was not related to previous questionnaire items on anxiety reactions, but this lack of relationship may have been the result of removing many highly anxious *S*s from the training class before the second questionnaire was given at the end of mock-tower training. The trainees removed from the class before the end of mock-tower jumping, usually because the trainee refused to jump from the mock tower, did have higher scores on the anxiety scale than other trainees. Many anxious *S*s, as inferred from the first questionnaire, refused to jump, but those anxious *S*s who did jump from the mock tower achieved satisfactory jump performance as quickly as other *S*s and rated themselves no more afraid during mock-tower training than nonanxious *S*s. This all-or-none relationship between anxiety and performance may have some implications for other studies of performance under stress.

SUMMARY

This research reported the use of a self-rating scale in a fearful situation, and inferred that a tendency to admit fear on the rating scale was an aspect of susceptibility to stress. Airborne trainees were asked to mark on a self-rating scale the amount of fear they felt during mock-tower jumping. One group of trainees rated fear in the mock tower just *prior* to jumping and the second group completed the rating scales *after* they had finished mock-tower training. For both groups performance was related to the self rating of fear. Those who subsequently passed the Airborne course estimated that they were less afraid than those who later failed it. Of the trainees who passed the Airborne course, those who achieved correct jump technique early in mock-tower training marked lower fear ratings than those who did not learn correct jump technique until late in training.

The group that was given the questionnaire after mock-tower training was asked other questions on fear reactions. Those who rated themselves as high in fear also answered that they were more "scared" during training and said that they had more physiological reactions to mock-tower jumping than did those who rated themselves as low in fear. From a questionnaire given prior to the start of Airborne training, an attempt was made to find the attitudes related to a high self rating of fear for learning to jump from the mock tower. Trainees who admitted high fear on the self-rating scale tended to be those who, on the first questionnaire, showed negative answers on questions about danger, said they were more worried about injury in Airborne training or in combat, and admitted less confidence in their ability to perform adequately in combat or in parachute jumping.

The study appeared to show the reliability and validity of a self rating on fear. The finding that admission of fear was related to previous confidence about stressful situations led to the suggestion that questions specific to stress might be fruitful in investigating the correlates of susceptibility to stress.

REFERENCES

1. AX, A. F. The physiological differentiation between fear and anger in humans. *Psychosom. Med.*, 1953, **15**, 433–442.
2. BROWN, J. S., & FARBER, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, **48**, 465–495.
3. CARLSON, V. R., & LAZARUS, R. S. A repetition of Meyer Williams' study of intellectual control and associated Rorschach factors. *J. consult. Psychol.*, 1953, **17**, 247–253.
4. ERIKSEN, C. W., LAZARUS, R. S., & STRANGE, J. R. Psychological stress and its personality correlates. Part II. The Rorschach test and other measures. *J. Pers.*, 1952, **20**, 277–286.
5. FLANAGAN, J. C. *The aviation psychology program in the Army Air Forces*. Washington: U. S. Government Printing Office, 1948. (AAF Aviat. Psychol. Program Res. Rep. 1.)
6. HAMMOCK, J. C., & PRINCE, A. L. Rifle marksmanship as a function of manifest anxiety and situational stress. *Amer. Psychologist*, 1954, **9**, 389–390. (Abstract)
7. JOHNSON, I., & JOHNSON, ELECTA. South Seas' incredible land divers. *National Geographic*, 1955, **107**, 77–92.
8. LAZARUS, R. S., DEESE, J., & OSLER, SONIA F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.
9. MELTON, A. W. *Apparatus tests*. Washington: U. S. Government Printing Office, 1947. (AAF Aviat. Psychol. Program Res. Rep. No. 4.)
10. OSLER, S. F. Intellectual performance as a function of two types of psychological stress. *J. exp. Psychol.*, 1954, **47**, 115–121.
11. OSS ASSESSMENT STAFF. *Assessment of men*. New York: Rinehart, 1948.
12. SHAFFER, L. F. Fear and courage in aerial combat. *J. consult. Psychol.*, 1947, **11**, 137–143.
13. STOPOL, M. S. Rorschach performance in relation to two types of stress. *J. consult. Psychol.*, 1954, **18**, 11–15.
14. STOUFFER, S. A., LUMSDAINE, A. A., LUMSDAINE, M. H., WILLIAMS, R. N., JR., SMITH, M. B., JANIS, I. L., STAR, S. A., COTTRELL, L. S., JR. *The American soldier*. Vol. II. *Combat and its aftermath*. Princeton: Princeton Univer. Press, 1949.
15. WESTROPE, MARTHA. Rorschach indices, manifest anxiety and performance under stress. *J. abnorm. soc. Psychol.*, 1953, **48**, 515–524.
16. WILLIAMS, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21–29.

Received February 25, 1955.

A FACTORIAL STUDY OF ADJUSTMENT TO STRESS[1]

WAYNE H. HOLTZMAN AND M. E. BITTERMAN

The University of Texas

A RECENT review of the literature on "anxiety and reactions to stress" demonstrated "an urgent need for research of a comparative and integrative type in which measurements of many different kinds are made on the same group of subjects and their interrelations systematically studied" (9, p. 20). The major features of such an investigation, reported in detail elsewhere (10), are here briefly summarized. Its purpose was to search for common factors in a wide variety of measures (ranging from sociometric ratings of personality to biochemical determinations of urinary constituents) which were assumed, on the basis of previous experimental work, to bear on the prediction of adjustment to stress.

METHOD

Subjects

The *S*s of the experiment were 135 cadets in the Air ROTC unit at the University of Texas.[2] Through the cooperation of the Department of Air Science and Tactics, the entire junior class as well as the senior cadet officers participated in the work. Participation was regarded as part of the regular AROTC program for the spring semester of 1952. The *S*s were informed that they were taking part in a study which was part of a larger Air Force research project on the selection of flying personnel, but that the results would in no way influence their status in the service.

Tests and Variables

There were two sessions of individual tests and three sessions of group tests which were scheduled during the first three months of the semester. Several months later, additional data were collected in the summer camps to which most of the men were sent. The group of tests selected for study was widely diversified. Represented in the battery were ratings of adjustment by self, peers, and instructors, objective and projective personality tests, stress tests, a galvanic conditioning situation, perceptual tests, and a urine analysis. A total of 101 different variables were derived from these procedures, of which 41 variables were selected for inclusion in the final correlation matrix (Table 1). To compute the intercorrelations among all 101 measures would have been most wasteful in view of the small number of *S*s and the possibility of spurious relationships.

Ratings. Three sets of ratings were made—sociometric ratings, ratings by instructors, and summer camp nominations. For the purpose of sociometric ratings, the 135 *S*s were divided into subgroups of 15 to 24 men who had attended classes and drill meetings together for at least six months. The task of each *S* was to rank himself and those men in his subgroup whom he felt he knew well enough on each of five traits: Adjustment ("handles personal problems well, rarely shows signs of excessive worry or anxiety, seems basically happy"), Officer Aptitude ("all-round Air Force Officer"), Crack-Up ("most likely to 'crack up' under the emotional strain of combat"), Bail-Out ("If you had to bail out over enemy territory, which of these men would you have the most confidence in to lead the group back to safety?"), and Familiarity ("Place the individual at the top whom you feel you know better than any of the others."). In addition, 10 Air ROTC instructors ranked all the men on two traits, Officer Aptitude and Crack-Up. Finally, summer camp groups were asked to name the men considered to be in the top third of each flight with respect to Officer Aptitude and the men considered to be in the lowest third of each flight. Since the great majority of men in each flight were from other schools, these judgments provided a largely independent measure of the trait in question.

Nine variables derived from these procedures were included in the final correlation matrix: ratings on Adjustment by peers and by self; ratings on Crack-Up, Bail-Out, and Familiarity by peers; an acquaintance index (derived from the number of men choosing to rate each *S*); and ratings of Officer Aptitude by instructors and by peers at the University, and by peers at summer camp. All variables involving pooled judgments of each *S* by his peers were selected because of their high reliability and their high face validity. Of the variables derived from self ratings, that for Adjustment was retained because of its significance in previous studies (4, 8). Of the two sets of instructor ratings, those for Officer Aptitude were retained because they semed easier to make. The summer camp nominations were included because of their high face validity, and the acquaintance index was included on an intuitive basis since no previous work had been done with it.

[1] This research was conducted under terms of contract AF 33(038)-13887 between USAF and The University of Texas in coordination with the Department of Clinical Psychology, USAF School of Aviation Medicine, Randolph Field, Texas. Special acknowledgment is due Dr. Saul B. Selis for his numerous valuable suggestions regarding the design and execution of this project. We are also indebted to J. Krauskopf, M. Varley, A. Kubala, J. Wodinsky, R. Duryea, C. Wood, E. Wortz, and C. Elam, graduate students at the University of Texas, for assistance in the collection and analysis of data. The Department of Biometrics, USAF School of Aviation Medicine, computed the correlations for the final intercorrleation matrix. A brief report of this research was presented in a paper at the 1954 meetings of the American Psychological Association, September 5, New York City.

[2] For most of the variables in this study, the actual *N* is somewhat less than 135 due to illness, failure to appear for testing, and resignation from the program.

Personality tests. The SAM Group Ink-Blot Test, a modification of the Rorschach, was included in the battery on the basis of promising preliminary results with the test as a device for the psychiatric screening of aviation cadets (16). It contributed four variables to the final matrix—the sum of plus signs, the sum of minus signs, the algebraic sum of plus and minus signs, and the C score. The first, third, and fourth scores were presumed to be positively related to adjustment, and the second score negatively related.[1] A single score derived from the Saslow Screening Test (15), previously found to be correlated with psychiatric ratings of anxiety proneness in Air Force personnel (19), also was included. The MMPI was administered to the *Ss*, and of the 17 scores derived only four were retained: *Mf*, *Sc*, *A*, and *IR*. Taylor's A scale (17), highly saturated with items from *Hy*, *Hs*, and *D*, was chosen to represent the neurotic triad, *Sc* was chosen to represent the psychotic triad, and *Mf* to represent the third component revealed by recent factorial studies of the test. The fourth score, Welsh's Internalization Ratio (20), was included on the basis of promising results obtained in the differentiation of various psychiatric groups.

Stress tests. Three situations were employed in which the performance of *S* under conditions of stress was measured. The first was the McKinney Reporting Test, a paper-and-pencil test suitable for group administration which involves time pressures and suggestions of substandard performance (11). This test contributed three scores to the final correlation matrix (poststress performance minus prestress performance, stress minus prestress, and late stress minus early stress). The second test in this group involved the measurement of steadiness by means of an apparatus employed in an earlier study (2). Measurements were made under conditions of loud noise and painful cutaneous shock, as well as under appropriately distributed control conditions. This test contributed five measures (reflecting basal level, effect of stress, and aspects of recovery). In the third stress test, *Ss* were required to trace a pattern under conditions of a 180° rotation of the visual field (produced by the so-called SAM Pseudoscope) and instruction to achieve maximum speed and accuracy. The *S*'s task was complicated by flashing red illumination and by vibration of the tracing surface. Two scores from this test were used, one reflecting speed (total time) and the other accuracy (error time).

Conditioned galvanic skin response. Measures of autonomic conditioning were made under conditions closely resembling those employed in previous studies (2, 3). In the present experiment, however, a discrimination procedure was employed. The conditioned stimuli were two small lamps, one red and one white, set side by side in a panel. There were six adaptation trials during which the lamps were illuminated, singly and in combination, without reinforcement. Then there were 20 training trials. On 10 of these trials both lamps flashed simultaneously and were followed five seconds later by a brief shock. On the other 10 trials the red or white lamp flashed alone and was not followed by shock. The interval between trials was 24 seconds.

From the GSR data, 22 scores were derived which were designed to reflect various aspects of response to the conditioning procedure—reactivity to light and shock, adaptation, conditioning, differentiation, and so forth. These measures were separately factor-analyzed to find a smaller number of variables in terms of which to represent *S*'s performance. As a result, six variables were chosen for inclusion in the final correlation matrix —L1 (response to light on Trial 1) and S7 (response to shock on Trial 7) as measures of reactivity, L25 as a measure of conditioning, L26 as a measure of extinction, L1 − L7 (the difference between the response to light on Trials 1 and 7) as a measure of adaptation to light and shock, and L18 − L20 as a measure of differentiation.

Perceptual tests. Three perceptual situations were employed in the experiment. The first involved measurements of size constancy, using a variable and a standard equilateral triangle placed 12 and 6 feet, respectively, from *S*, and a comparable measure of shape constancy (5). Four scores were derived for the final matrix—two measures of constancy (size and shape) and two corresponding measures of anticipation (based on the difference between the results of ascending and descending determinations). The interest in constancy measures stemmed from evidence suggesting that they may have diagnostic value for certain neurotic and psychotic categories (13, 14). The second perceptual situation was a tachistoscopic one, in which thresholds for taboo and neutral control words, as well as thresholds for the perception of incongruity, were studied with the Gerbrands apparatus. Because of the unreliability of these measures, none was included in the final matrix. The third perceptual situation was designed to reflect the relative dominance of visual and kinesthetic cues in the discrimination of the gravitational vertical. It was modeled after the situations employed by Witkin (21). The *S* sat in a chair inclined 22° to the right, looking into a cubical enclosure tilted 28° to the left, and attempted to adjust a rod (which moved independently of the surround) to the gravitational vertical. The test provided a single score, the mean absolute deviation from the vertical in a series of eight settings.

Biochemical variables. In a variety of recent studies, normal and psychiatric groups have been differentiated in terms of adrenocortical response to stress (9), and the present study was originally designed to include measurements of the effect of ACTH on the output of 17 ketosteroids and on blood eosinophils. Since University authorities would not permit these measurements, a more limited objective was adopted. Early-morning samples of urine were analyzed for 22 biochemical constituents reflecting metabolic activity. Only two measures, uric acid and glycine, were included in the final matrix. These measures were selected on the basis of the normality of their frequency distributions, their known day-to-day stability, and their interrelations with the other biochemical variables as determined in a separate analysis.

RESULTS

Each of the 41 selected variables was coded and punched on IBM cards. Where the

[1] Scores for the SAM Ink-Blot Test were based upon experimental work subsequent to the earlier report by Sells (16), although the method of administration was identical to that which he used.

distribution for a given variable was sharply skewed, it was normalized prior to coding. The obtained correlation matrix is presented in Table 1.

To prevent serious distortion of factorial structure because of the large number of low intertest correlations, cluster analyses were made of each set of intratest correlations, and only those variables from a given test which appeared relatively unique were retained for further study. By Thurstone's centroid method (18), seven significant factors were extracted from the intercorrelations of the 25 selected variables. The reference frame was rotated first without knowledge of the identity of any of the variables, using simple structure as the primary criterion for location of axes (18). Since the communalities of the variables were generally rather low, however, it was difficult to obtain a satisfactory solution without considerable adjustment of the reference axes after careful inspection of the nature of the variables. The rotated orthogonal factor loadings are given in Table 2.

The variables which define the content of each of these seven factors, with interpretation and discussion, are given below.

I. Defining variables for the first factor are as follows:

Variable	*Loading*
9. Officer Aptitude, Peer	.80
12. Officer Aptitude, Camp Nomination	−.65
11. Officer Aptitude, Instructor	.58
7. Familiarity	.42

This is the clearest factor in the group. Since Officer Aptitude, Peer, has a correlation of .72 with Adjustment, Peer, (Table 1) it may be concluded that manifestations of good social adjustment are a prerequisite to being considered by one's peers as possessing a high degree of officer aptitude. The appreciable loading of Familiarity on Factor I indicates that individuals who have high officer aptitude are generally those who are fairly well known by their associates.

II. Defining variables for the second factor are as follows:

Variable	*Loading*
49. SAM Pseudoscope: Error Time	.61
48. SAM Pseudoscope: Total Time	.53
72. Shape Constancy: Anticipation	.43
70. Size Constancy: Anticipation	.41

This factor cuts across several tests of perceptual and perceptual-motor functioning. High scores on the SAM Pseudoscope are due to the inability of *S* to trace a pattern quickly and accurately under conditions of perceptual-motor conflict. The overwhelming majority of *S*s showed positive anticipatory error scores on the two constancy tests. Variables 70 and 72 measure the tendency for *S*'s judgments to be distorted by directional influences. Unfortunately, the functional significance of these relationships is unclear, although it may be said that high scores on all four variables reflect inability to cope effectively with conflicting perceptual cues.

Of some interest is the fact that the third perceptual test included in the factor analysis, Vertical Orientation (Variable 73), has a loading of −.13 on Factor II. Witkin and coworkers (21) obtained significant correlations between tests of space orientation similar to the one herein used and other tests involving perceptual-motor conflict and contradictory visual cues, as well as certain personality measures. They concluded that individuals tend to be self-consistent in their functioning under quite different conditions—i.e., behave consistently with respect to degree of dependence upon the visual field regardless of the particular task. Since "field dependence" on Vertical Orientation would indicate inability to handle conflicting cues, Variable 73 should have a high positive loading, rather than a low negative one, on Factor II.

III. Defining variables for the third factor are as follows:

Variable	*Loading*
36. McKinney: Stress–Prestress	.67
35. McKinney: Poststress–Prestress	.57
17. Saslow Screening Test	.35

This factor is defined primarily by the McKinney Reporting Test. The appreciable loading on the Saslow Screening Test is of interest, however, since this short self inventory was given immediately after the somewhat frustrating test of McKinney. It may be that responses to the items of Saslow's inventory reflected in some degree the extent to which *S* was disturbed by the McKinney test. Since high scores on Variables 35 and 36 indicate deterioration of performance under stress, this factor apparently represents one aspect of inadequate behavior under stress.

TAB

INTERCORRELATIONS

Variable No.	Variable Name	Intercorrelations														
1	Adjustment, peer															
*2	Adjustment, self	08														
3	Crack-Up, peer	−66	−17													
6	Bail-out	−58	−06	55												
*7	Familiarity	37	00	−30	−19											
8	Acquaintance index	−13	−22	12	−08	−16										
*9	Officer aptitude, peer	72	09	−62	−54	28	−15									
*11	Officer aptitude, instructor	40	−02	−28	−30	28	11	50								
*12	Officer aptitude, camp nominations	41	03	−45	−28	09	−07	55	16							
13	SAM Ink-blot: plus	22	−11	−12	−02	27	03	18	06	32						
14	SAM Ink-blot: minus	−06	02	13	03	−19	−01	−17	03	−31	−89					
*13	SAM Ink-blot: sum	07	−07	−13	−03	24	02	18	01	32	97	−97				
16	SAM Ink-blot: C	03	−10	−12	−01	26	03	15	07	24	89	−88	87			
*17	Saslow Screening Test	06	28	−07	−02	−11	−02	04	−06	24	−12	07	−09	−15		
*26	MMPI: *Mf*	−03	03	07	01	−01	03	−04	07	09	10	−14	12	06	12	
*29	MMPI: *Sc*	08	−02	−02	05	−03	09	07	13	03	−06	06	−06	−09	12	27
*32	MMPI: *A*	22	24	−12	−20	01	12	19	12	23	05	−02	03	07	41	25
*34	MMPI: *IR*	04	35	−08	00	07	−09	−02	−09	13	−01	−01	00	−02	08	11
*35	McKinney: Post-Prestress	−02	20	05	−03	−05	00	00	−21	14	07	−14	11	09	14	08
*36	McKinney: Stress-Prestress	−06	10	04	10	−01	00	04	−02	36	06	−08	07	08	24	04
37	McKinney: Late-Early Stress	04	02	−07	−02	−12	−04	03	03	14	−07	00	−04	−14	01	07
*38	Steadiness: basal level	16	18	−11	04	11	−04	09	15	27	07	−01	04	14	14	−02
40	Steadiness: Pre-Poststress	07	26	−14	−02	07	06	−01	02	04	05	−01	03	07	02	−12
*42	Steadiness: Early-Prestress	00	−05	07	−02	−04	06	02	−03	19	04	−03	04	04	05	02
44	Steadiness: Stress Accommodation	−04	−14	01	02	−06	15	−06	−20	24	07	−05	06	08	06	08
47	Steadiness: Total recovery	−03	02	04	01	−01	00	05	07	11	−09	06	−08	−10	04	−03
*48	SAM Pseudoscope: total time	01	00	19	01	−31	16	−12	−09	08	−03	01	−02	−10	01	−09
*49	SAM Pseudoscope: error time	06	07	−06	00	−18	−03	01	03	17	00	−01	00	00	08	−01
*50	GSR: L1	07	15	−05	−06	−06	05	11	21	11	−05	10	−08	−06	00	−12
52	GSR: L1–L7	03	25	08	03	03	01	04	09	03	−10	10	−10	−07	03	−07
54	GSR: S7	08	14	−03	−02	−02	03	11	21	10	05	−04	05	07	10	−04
*60	GSR: L18–L20	05	26	−10	03	07	−13	07	−02	−08	−04	−04	00	10	05	07
*62	GSR: L25	−02	07	03	−08	−10	11	06	14	−29	01	00	00	05	01	05
*63	GSR: L26	01	−18	−16	00	−12	13	−07	12	−06	−13	13	−13	−21	−03	−11
*69	Size constancy: total	01	−15	09	−06	00	13	02	−01	01	04	−03	03	07	−10	−08
*70	Size constancy: anticipation	09	02	01	−06	05	−07	06	08	−14	07	−07	07	11	04	06
71	Shape constancy: total	−02	−02	16	11	−03	04	−03	01	04	01	03	−01	−01	−04	15
*72	Shape constancy: anticipation	−05	−08	07	11	−11	07	−07	−06	01	−23	21	−23	−18	01	−14
*73	Vertical orientation	−04	00	−09	−15	09	04	00	01	08	−12	11	−12	−08	−04	−02
81	Uric acid	−09	03	−12	03	−25	03	05	−07	19	07	−16	12	13	00	15
85	Glycine	−06	06	03	05	−03	02	00	−01	00	10	−08	09	10	−15	22
	Variable Number	1	2	3	6	7	8	9	11	12	13	14	15	16	17	26

* Indicates variables included in the final factor analysis.

IV. Defining variables for the fourth factor are as follows:

Variable	*Loading*
62. GSR:L25	.77
63. GSR:L26	.61
50. GSR:L1	.32

Factor IV seems to reflect mainly individual differences in autonomic reactivity and conditioning to shock in the GSR test. It is interesting to note that none of the other tests in the factor matrix has an appreciable loading on Factor IV, although several of the relationships are suggestive (Table 2, Variables 11, 12, 42, and 69).

V. Defining variables for the fifth factor are as follows:

Variable	*Loading*
32. MMPI: *A*	.59
34. MMPI: *IR*	.47
2. Adjustment, Self	.46
26. MMPI: *Mf*	.39
17. Saslow Screening Test	.38
29. MMPI: *Sc*	.33

This factor is one of the most clearly differentiated in the entire analysis. Although all four of the scales from the MMPI have significant loadings on this factor, the loadings on two independent variables (2 and 17) de-

LE 1
OF MAJOR VARIABLES

Intercorrelations (*continued*)

23																									
03	22																								
09	11	13																							
−03	14	07	46																						
04	−06	21	13	11																					
22	06	12	05	−01	−03																				
04	−06	07	14	60	13	60																			
12	05	18	06	−07	−08	−03	−23																		
05	04	12	13	−04	−01	06	06	68																	
07	−04	18	09	01	01	09	04	77	47																
07	03	−01	−01	06	04	−02	−07	18	07	08															
20	−02	13	03	−09	01	23	11	19	04	19	40														
05	14	−01	05	00	−05	06	22	−07	−07	−01	−05	01													
−03	13	03	14	02	−09	06	17	−10	00	−02	−06	−02	76												
−10	09	00	−10	06	−10	02	13	01	06	07	01	−07	56	43											
−07	04	−04	06	06	−04	−02	09	−24	−17	−21	−21	−07	09	12	14										
−13	12	00	−03	07	−06	−16	−02	07	−05	05	04	−04	32	25	40	22									
−07	−08	−16	−12	02	−13	01	−13	31	18	22	−05	05	14	11	33	13	43								
−03	02	−03	03	−03	−19	−07	−08	04	07	−07	04	00	−06	−01	−15	04	−24	−10							
−04	01	01	−03	07	07	04	−01	02	−12	−01	25	22	−05	−07	−04	05	−09	−08	18						
04	−01	06	02	07	11	14	03	01	−07	02	12	−06	14	11	09	−06	−05	−02	13	02					
−02	00	−01	11	09	02	22	13	03	07	10	13	16	14	02	14	03	06	16	−03	27	05				
−11	07	07	−11	−02	−12	−03	05	−03	03	04	−19	−10	14	16	09	03	07	−11	17	−20	−08	−05			
19	07	13	−07	−08	00	−05	−14	12	09	−01	17	15	−09	02	02	00	12	−05	−04	00	−14	−13	15		
−10	01	18	−10	−06	−01	03	04	−09	−14	−05	−12	−03	−11	−04	−16	18	00	−01	09	06	05	−04	06	−08	
29	32	34	35	36	37	38	40	42	44	47	48	49	50	52	54	60	62	63	69	70	71	72	73	81	

signed to measure self-admitted anxieties and the high loading on Taylor's scale (Variable 32) suggest that Factor V is especially related to anxiety. A high score on any of the variables is considered to be indicative of maladjustment.

Factors VI and VII appear to be mainly residual factors of little theoretical interest. Only Variables 2, 42, and 60 have moderate loadings on the sixth factor. Factor VII accounts for most of the common variance in Variable 15, although a significant negative loading does appear for one of the GSR measures (Variable 50). In neither of these residual factors does a pattern emerge clearly enough to permit any generalizations.

DISCUSSION

One can only be impressed with the generally low degree of intercorrelations obtaining in this study among the large number of variables which, at one time or another, have been classed together because of presumed relationships to such personality traits as adjustment, anxiety, or emotional stability and control. Most of the sizable correlations are among variables derived from the same or similar tests. The only variables with loadings greater

TABLE 2

ROTATED ORTHOGONAL FACTOR LOADINGS*

Variable Number	Factors							h^2
	I	II	III	IV	V	VI	VII	
2	−01	−03	23	04	46	−40	04	43
7	42	−25	−19	−09	04	−08	16	31
9	80	06	−06	01	01	−13	03	66
11	58	15	−33	21	03	−15	−11	54
12	65	11	25	−21	17	23	05	62
15	30	−08	00	−10	00	00	49	34
17	09	04	35	03	38	01	−09	29
26	05	−12	02	02	39	18	15	22
29	11	19	−09	−04	33	25	−13	25
32	19	−03	16	11	59	09	−16	45
34	00	−05	09	−01	47	05	17	26
35	07	03	57	−01	14	16	07	38
36	23	−01	67	07	−02	12	03	52
38	24	28	03	−11	25	−06	03	22
42	01	21	−10	21	12	43	23	35
48	−19	53	05	−09	−02	11	12	35
49	−04	61	−06	−10	21	04	12	44
50	08	09	−01	32	−03	−18	−35	27
60	−01	−10	13	17	09	−45	01	26
62	−08	−03	−01	77	05	−17	07	64
63	−01	18	−07	61	−21	17	05	48
69	05	00	−06	−21	−13	15	−10	10
70	−04	41	−01	−13	08	−15	15	24
72	−07	43	13	11	−03	05	−21	27
73	07	−26	−13	−02	16	01	−25	18

* Decimal points have been omitted.

than .30 on Factor I were those obtained from the personality ratings. Factor II was defined entirely by perceptual variables. None of the variables other than those from the GSR test had appreciable loadings on Factor IV. Each of the remaining four factors is represented in relatively diverse tests, but these transverse effects must be regarded as minor when the entire set of measures is considered as a whole.

Consistent with these results are those of a study by Parsons, Phillips, and Lane (12), employing apparatus markedly similar to the present steadiness test. From the assumption of a general factor of resistance to stressful conditions, it was inferred that performances on the same psychomotor task under two different stressful conditions should be related. No significant correlations were found, suggesting that different variables were operating in the two situations even though the task was identical. If such specificity is present between two such seemingly similar measures as those obtained by Parsons, it is understandable how so little communality was discovered among the many variables employed in the present experiment. It may well be that no general factor exists.

Convincing evidence of the relative specificity of the variables selected for study is provided by the generally low communalities. The average communality of the 25 variables is only .36, leaving nearly two-thirds of the test variance unexplained. While an appreciable part of this specific variance is probably due to the low reliabilities of some of the measures used, there still remains considerable variance that is stable and relatively unique. How much of this apparently specific variance can be attributed to common factors untouched by the present intercorrelations remains to be determined.

SUMMARY

The purpose of this study was to search for common factors in a variety of measures believed on the basis of previous research to be of value in the prediction of adjustment to stress. The variables selected for analysis were derived from ratings of personality and officer aptitude, objective and projective personality tests, measures of performance in stressful situations, the conditioning of the galvanic skin response, perceptual tests, and the analysis of urinary components. The *S*s were 135 Air ROTC cadets at the University of Texas.

Analysis of the intercorrelation matrix yielded seven factors: Factor I was defined chiefly by measures derived from the ratings, II by perceptual measures, III by variables derived from one of the stress tests, IV by conditioning measures, and V by MMPI measures; the meaning of VI was in doubt, and VII appeared to be mainly a residual factor. In general, the degree of relationship among the measures derived from the different types of tests was very low, although in some cases there were statistically significant and meaningful correlations between the ratings, the personality tests, and the laboratory variables.

REFERENCES

1. BERRY, HELEN K. Variations in urinary excretion patterns in a Texas population. *Amer. J. phys. Anthrop.*, 1953, **11**, 559–575.
2. BITTERMAN, M. E., & HOLTZMAN, W. H. Conditioning of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.
3. BITTERMAN, M. E., KRAUSKOPF, J., & HOLTZMAN,

W. H. The galvanic skin response following artificial reduction of the basal resistance. *J. comp. physiol. Psychol.*, 1954, **47**, 230–234.

4. CALVIN, A. D., & HOLTZMAN, W. H. Adjustment and the discrepancy between self concept and inferred self. *J. consult. Psychol.*, 1953, **17**, 39–44.
5. GOTTHEIL, E., & BITTERMAN, M. E. The measurement of shape-constancy. *Amer. J. Psychol.*, 1951, **64**, 406–408.
6. HOLTZMAN, W. H. Validation studies of the Rorschach test: Shyness and gregariousness in the normal superior adult. *J. clin. Psychol.*, 1950, **6**, 343–347.
7. HOLTZMAN, W. H. Validation studies of the Rorschach test: Impulsiveness in the normal superior adult. *J. clin. Psychol.*, 1950, **6**, 348–351.
8. HOLTZMAN, W. H. Adjustment and leadership: A study of the Rorschach test. *J. soc. Psychol.*, 1952, **36**, 179–189.
9. HOLTZMAN, W. H., & BITTERMAN, M. E. Psychiatric screening of flying personnel. VI. Anxiety and reactions to stress. *USAF Sch. Aviat. Med.* 1952, Proj. No. 21-37-002, Rep. No. 3.
10. HOLTZMAN, W. H., & BITTERMAN, M. E. Adaptability screening of flying personnel. The prediction of adjustment to stress: A factorial study of selected measures. *USAF Sch. Aviat. Med.*, to be published.
11. MCKINNEY, F., STROTHER, G. B., HINES, R. R., & ALLEE, R. A. Experimental frustration in a group test situation. *J. abnorm. soc. Psychol.*, 1951, **46**, 316–323.
12. PARSONS, O. A., PHILLIPS, L., & LANE, J. E. Performance on the same psychomotor task under different stressful conditions. *J. Psychol.*, 1954, **38**, 457–466.
13. RAUSH, H. L. Perceptual constancy in schizophrenia: I. Size constancy. *J. Pers.*, 1952, **21**, 176–188.
14. SANDERS, R., & PACHT, A. R. Perceptual size constancy of known clinical groups. *J. consult. Psychol.*, 1952, **16**, 440–444.
15. SASLOW, G., COUNTS, R., & DUBOIS, P. Evaluation of a new psychiatric screening test. *Psychosom. Med.*, 1951, **13**, 242–253.
16. SELLS, S. B., FRESE, F. J., & LANCASTER, W. H. Research on the psychiatric selection of flying personnel. II. Progress on development of SAM Group Ink-Blot Test. *USAF Sch. Aviat. Med.* 1952, Proj. No. 21-37-002, Rep. No. 2.
17. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
18. THURSTONE, L. L. *Multiple-factor analysis.* Chicago: Univer. of Chicago Press, 1947.
19. ULETT, G. A., GLESER, GOLDINE, LAWLER, A., & WINOKUR, G. Psychiatric screening of flying personnel. IV. An experimental investigation of development of an EEG index of anxiety tolerance by means of photic stimulation—its validation by psychological and psychiatric criteria. *USAF Sch. Aviat. Med.* 1952, Proj. No. 21-37-002, Rep. No. 4.
20. WELSH, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, **16**, 65–72.
21. WITKIN, H. A., LEWIS, HELEN B., HERTZMAN, M., MACHOVER, KAREN, MEISSNER, PEARL B., & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

Received March 2, 1955.

MOTIVATION AND PERFORMANCE IN SCHIZOPHRENIA[1]

BERTRAM D. COHEN[2]
State University of Iowa

THE impairment of performance implied by the term *psychological deficit* (4) is a consistent correlate of schizophrenia. The evidence for this relation between diagnosis and performance is extensive (*e.g.*, 3, 4, 6, 8, 9, 12, 14). The present paper is concerned with the role of motivation as a determinant of deficit in schizophrenia.

In previous studies of motivation and performance in schizophrenia test scores have been correlated with behavioral ratings defined as indices of motivation. For a group of schizophrenic patients, Wittman (15) found high positive correlations (median $r = +.75$) between cooperativeness ratings and scores on a battery of reasoning and memory tests. The corresponding correlations for a group of paretic patients were low, approachng zero. Wittman interpreted her results as indicating that motivational factors were responsible for the deficit shown by the schizophrenic group. She attributed the deficit shown by the paretic patients to nonmotivational factors, since their test scores were unrelated to their cooperation ratings. Shakow (14) summarized results of cooperation-performance correlations for a series of studies of motor function, reaction time, and general intelligence in schizophrenia. The median r reported was $+.56$.

These results are complemented by clinical observations that schizophrenic patients' efficiency can be raised with favorable changes in examiner-patient relationships (2). There remains, however, a lack of data relating task performance in schizophrenia to experimentally manipulated motivational variables.

THE PROBLEM

If impaired motivation is a significant determinant of the performance deficit shown by schizophrenic patients, then an experimentally produced increase in motivation should improve performance. In normal *S*s, however, relatively little improvement in performance can be expected with experimentally increased motivation. This latter prediction rests on the assumption that "rapport" alone (appropriate instructions and encouragement) yields optimal motivation in normal individuals. That is, they work near their limit of proficiency under ordinary testing conditions, leaving relatively little room for improvement with experimentally increased motivation. The present study was designed to test the hypothesis that an experimentally produced increase in motivation will occasion greater improvement in the performance of schizophrenic patients than in normal individuals.

The task set for the *S*s was the learning of a series of motor responses to successively presented stimuli. The task involved rote memorization and visual-motor coordination. As each stimulus was presented, *S*s were required to make either a right or left movement of a handle. Mastery of the task was attained when the series of movements was completed without error.

Schizophrenic and normal *S*s were tested under ordinary rapport conditions and under a condition in which an increment in motivation was experimentally induced by the use of electric shock. The shock began concurrently with the visual stimulus. The cessation of shock was contingent on the making of a correct response.

EXPERIMENTAL METHOD

Subjects

The schizophrenic group was composed of 38 patients selected from three mental hospitals with large custodial facilities.[3] The patients presented diagnosti-

[1] This paper is adapted from a thesis submitted in partial fulfillment of the requirements for the Ph.D. degree at the State University of Iowa. The author wishes to acknowledge the advice and criticism of Drs. John R. Knott and Judson S. Brown, under whose sponsorship the study was carried out. Special thanks are also due Drs. Paul E. Huston and Rita Senf for their many helpful suggestions.

[2] Now with the Dept. of Neuropsychiatry, Iowa City VA Hospital and Psychology Dept., State University of Iowa.

[3] Of the patients used, 23 were from the Mt. Pleasant Mental Health Institute, Mt. Pleasant, Iowa; 11 were from Elgin State Hospital, Elgin, Illinois; and 4 were from the VA Hospital, Knoxville, Iowa.

cally clear symptoms of schizophrenia, as judged from case records and physicians' recommendations. The mean age was 35.3 years, with a median of 36.5 and a range from 20 to 46 years. The mean number of grades completed was 11.2, with a median of 12 and a range from 7 to 17. There were 29 men and 9 women. Subtype composition of the group was as follows: 15 paranoid, 5 hebephrenic, 4 simple, 3 catatonic, 1 mixed, and 10 unclassified.

The 32 normal *S*s were obtained from two sources: nonpsychiatric patients from the State University of Iowa Hospitals, and volunteers secured through the local Methodist Church young adults' groups. The mean age was 32.6 years, with a median of 32.0 and a range from 22 to 45. The mean number of grades completed was 11.9, with a median of 12 and a range from 7 to 17. There were 24 men and 8 women.

Apparatus

The stimulus presentation panel and the *S*'s reaction handle were housed in a portable wooden cabinet. In the center of the front face of the cabinet was a ground glass screen 2 inches wide by 20 inches high. The reaction handle was on a platform directly below the stimulus presentation panel. It could be moved laterally through a 12-inch groove in the center of the platform. The starting position of the reaction handle was always at the midpoint of the groove. Although a full 6-inch movement was required for the correct reactions, a movement of only 1½ inches in the wrong direction activated an error counter.

In a vertical row behind the screen, ten 6.3-volt dial lamps were spaced at intervals of 1⅞ inches. Each lamp was separated from the adjacent ones by metal partitions. A cardboard strip could be inserted between the lamps and the screen. Any one strip contained ten identical forms, cut to correspond to the position of the lamps. This permitted the projection of distinctively shaped visual stimuli on the screen.

The presentation of successive stimuli from top to bottom could be controlled either by *E* or by *S*'s reactions. Each trial was initiated by *E*'s turning on the top lamp. The *S* then moved the reaction handle either to the left or right. When he made the correct movement, the top lamp went out and, after a three second delay, the next lamp in the series came on automatically. The correct series of movements could be set in advance by control switches.

A wristwatch-type band, containing two electrodes, was fastened around *S*'s left wrist. Shock was provided by a battery-powered Harvard inductorium. When administered, shock began automatically with the lighting of a lamp and ended when the lamp went out following a correct movement. Shock could also be controlled by *E* independently of *S*'s reactions.

Procedure

All *S*s, regardless of later group assignment, were first put through a shock adjustment procedure. For each *S* the shock level was adjusted midway between an "uncomfortable" and a "painful" level. The electrodes remained in position throughout the testing, whether or not shock was administered. Any patient who became too agitated or objected strenuously was dropped from the experiment.

A practice pattern was used to illustrate the task. The *E* demonstrated the correct sequence of movements, which was: LLRRLRLLRR. Then *S* performed the same task until he showed apparently adequate comprehension of directions. Praising, urging, and instruction were freely used.

A second stimulus strip, Pattern I, was next inserted. The correct sequence was: RLRRRLRLLR. The *S*s were instructed to learn the correct pattern. No demonstration was given, so the first trial was exploratory. If *S* did not reach the criterion of one perfect sequence by the twentieth trial, he was eliminated from the study. Such failures to learn seemed chiefly due to extreme anxiety, extreme suspicion of *E* or the apparatus, or severe psychotic disorganization. Patients retained were therefore relatively well-preserved and cooperative.

At this stage the schizophrenic and normal groups were each divided into two procedural subgroups, Shock and Rapport. Thus there were four groups: *Group PR: Psychotic Ss* who learned the experimental patterns (Patterns II and III; see below) under the nonshock or *rapport* condition; *Group PS: Psychotic Ss* learning Patterns II and III under the *shock* condition; *Group NR: Normal Ss* learning Patterns II and III under the *rapport* condition; and *Group NS: Normal Ss* learning Patterns II and III under the *shock* condition.

The *S*s in Group PR were matched with those in Group PS on the basis of number of trials to master Pattern I, with a discrepancy of no more than one trial permitted. Groups NR and NS were similarly matched.

Instructions for Patterns II and III were similar to those for Pattern I. Praise and encouragement were again used freely for all groups. The only difference in procedure was that Groups PS and NS were given shock during performance. Instructions prior to Patterns II and III clearly indicated to *S*s in the shock groups that shock would start with the onset of each light and cease with the completion of the correct movement. Brief shocks were administered before beginning Pattern II and, where necessary, adjustments were made in shock level. During performance, readjustments, either upward or downward, were sometimes made to keep the shock "uncomfortable" but not to produce obvious agitation.

The correct sequence for Pattern II was: RLLRRLRRL. Learning was continued to the criterion of one perfect trial. If *S* did not reach this criterion by the twentieth trial, performance was terminated. In either event, Pattern III was administered next. Here the sequence was: LRRLRLLLRL. Performance was continued to the same criterion as for Pattern II.

RESULTS

The measures to be considered are the number of trials and the number of errors for each pattern. Table 1 presents the results for Pattern I, before the normal and schizophrenic groups were divided into Shock and Rapport subgroups. On both measures the schizophrenic

TABLE 1

COMPARISON OF NORMALS AND SCHIZOPHRENICS ON PATTERN I

	Normals (N = 32)		Schizophrenics (N = 38)		t
	Mean	SD	Mean	SD	
Trials	5.6	3.7	9.6	5.0	3.61
Errors	11.3	11.6	24.3	16.6	3.69

TABLE 2

COMPARISONS OF SHOCK AND RAPPORT GROUPS OF NORMALS AND OF SCHIZOPHRENICS SEPARATELY

	Normal Groups					Schizophrenic Groups				
	NR (N = 16)		NS (N = 16)			PR (N = 19)		PS (N = 19)		
	Mean	SD	Mean	SD	t*	Mean	SD	Mean	SD	t
Pattern II										
Trials	4.7	2.8	3.5	2.5	1.33	7.6	4.6	7.5	4.4	.03
Errors	7.9	7.6	4.8	4.8	1.40	17.3	13.5	19.2	16.7	.32
Pattern III										
Trials	3.4	2.2	4.3	3.5	.84	9.6	6.2	7.2	5.5	1.51
Errors	5.7	5.9	7.9	8.6	.75	23.5	16.5	15.2	14.0	1.91

* In this table t tests are for related measures; in other tables the t tests are for unrelated measures. All tests are for two-tailed hypotheses.

TABLE 3

COMPARISONS OF MEAN DIFFERENCES BETWEEN THE NORMAL GROUPS WITH THE MEAN DIFFERENCES BETWEEN THE SCHIZOPHRENIC GROUPS

	Mean Dif. NR - NS	Mean Dif. PR - PS	Mean Dif. between Difs. (NR-NS)–(PR-PS)	t
Pattern II				
Trials	+1.2	+0.1	+1.1	.65
Errors	+3.1	−1.9	+5.0	.32
Pattern III				
Trials	−0.9	+2.4	−3.3	1.68
Errors	−2.2	+8.3	−10.5	1.95

patients were significantly poorer ($p < .01$). Consequently one may accept the fact of deficit with assurance.

In Table 2 are the Pattern II and Pattern III data concerning the differences between the Rapport and Shock groups of normal and of schizophrenic groups separately. For the normal groups, the differences on Patterns II and III are not statistically significant for either the trials or errors measures. The addition of shock therefore failed to produce an appreciable effect on the performance of the normal *Ss*.

The data for Groups PR and PS (right half of Table 2) indicate no significant difference between these groups on Pattern II for either trials or errors. The difference between the groups fails to achieve significance for trials on Pattern III, but is moderately significant (between the 5 per cent and 10 per cent levels of confidence) for errors. It is of interest to note that for both trials and errors, the schizophrenic group learning Patterns II and III with the shock stimulation (Group PS) showed a continued decline in mean number of trials and errors from Pattern I through Pattern III. The schizophrenic rapport group (Group PR) also showed a decline in trials and errors from Pattern I to II, but a reversal of this trend from Pattern II to III.

To test the prediction regarding the differential effect of the shock on the performance of schizophrenic patients as compared with normals, the mean differences between Groups NR and NS were compared with the mean differences between Groups PR and PS on Patterns II and III. The prediction was that the shock stimulation would influence the performance of schizophrenic patients more favorably than the performance of normals. Thus it was expected that the difference NR minus NS would be smaller than the difference PR minus PS for the trials and errors measures on Patterns II and III.

The data are given in Table 3. While the differences were insignificant for Pattern II, the data indicate moderately reliable differences in both trials and errors on Pattern III. Group PS required on the average 2.4 fewer trials than Group PR, while Group NS required an average of 0.9 more trials than Group NR. The difference of 3.3 trials between these two mean differences is significant between the 5 per cent and 10 per cent levels of confidence and suggests that shock had a more favorable effect on the schizophrenic patients than on the normals. A similar result was obtained for the errors measure on Pattern III. A mean of 8.3 fewer errors was made by Group PS as compared with Group PR, while Group NS averaged 2.2 more errors than Group NR. The difference of 10.5 between these means is significant at approximately the 5 per cent level of confidence.

In summary, for Pattern II the data indicate no differential facilitative effects of the shock stimulation on the performance of the

schizophrenic patients as compared with the normals. On Pattern III, however, the data indicate that the shock did have a differential facilitative effect for Group PS as measured by trials and errors.

DISCUSSION

The results clearly confirm previous findings of deficit in learning tasks for schizophrenic patients (1, 5, 6). Motivational variables may not account for the deficit entirely, but the data support the hypothesis that variables of this class are responsible for some portion of the impairment. The predicted differential facilitative effect of the shock stimulation in schizophrenic patients was found, but only on the final pattern.

Inspection of Table 2 indicates that this separation of the schizophrenic groups was associated with an increase in trials and errors on Pattern III for Group PR relative to their own scores on Pattern II. No comparable increase in scores occurred for Group PS or for either normal group. It seems plausible to infer a process of dissipation of motivation in Group PR as the experimental session progressed. Since Group PS did not show a comparable decline in efficiency, it may be inferred that the shock stimulation sustained the motivation of these patients through Pattern III.

Kraepelin (7) inferred a similar process of motivational dissipation to account for his observation that, as compared with control *Ss*, patients with dementia praecox did not profit from interpolated rest intervals during performance on continuous addition tasks. He reasoned that the return to peak performance levels following rest occurred for the controls because their prerest drops were due to fatigue alone. Since the dementia praecox patients did not show similar postrest gains, he inferred a loss in "will-tension" to account for their performance drops. According to Hunt and Cofer, "Kraepelin appears to have based his judgment that dementia praecox manifests itself largely as a disturbance of volition" on a large number of observations of continuous work curves of this character. They note also that "with various types of tasks schizophrenics work efficiently for a brief period, but performance falls off at an abnormally high rate as they continue" (4, p. 998). The present pattern of results is consonant with these observations.

The results suggest two generalizations about the motivational component of deficit in schizophrenia. The first is that in a task situation schizophrenic patients are inadequately aroused by acquired motivational cues. The second is that this inadequacy can be partially compensated for by substituting primary motivation and reward (e.g., shock stimulation at onset of task stimulus and termination of shock with completion of correct response).

The findings of Sands and Rodnick (13) are pertinent to the first generalization. They measured the GSR to the word *ready* in schizophrenic patients and normals. The normals showed a response three times as great as the patients. It would be useful to know if in schizophrenic patients the GSR to shock stimulation is closer to normal than the GSR to acquired cues like the word *ready*. Data tending to confirm this have been reported by Paintal (10). His patient group, however, was composed of mixed psychotics rather than schizophrenic patients alone.

With regard to the second generalization there is, as yet, little evidence in the experimental literature. Some support for the idea may be taken from the studies of "multiple choice learning" in chronic schizophrenia by Peters (11). Peters indicated the usefulness of stimulating motivation in these patients by administering subshock doses of insulin and then using an appropriate sweet nutrient as reward for correct responses.

The trends in these investigations suggest a direction for research which promises a possible integration of the psychopathology of schizophrenia with general behavior theory.

SUMMARY

This study tested the effects of an experimentally produced increase in motivation on the performance of schizophrenic patients and normals. The task involved the learning of a series of motor responses to successively presented visual stimuli. The schizophrenic patients and normals were each divided into two groups, Shock and Rapport, matched on an initial test, Pattern I. The shock groups were then tested on two subsequent tasks, Patterns II and III, under conditions in which an increment in motivation was experimentally produced through the use of electric shock. The

rapport groups were similarly treated except that electric shock was omitted.

The results showed a differential facilitative effect of the shock stimulation for the schizophrenic patients on Pattern III, but not on Pattern II. The schizophrenic rapport group declined in efficiency from Pattern II to III, a result not seen in the shock group of patients or in either normal group.

These findings were interpreted in terms of an inferred process of motivational dissipation during performance in schizophrenia. Some further generalizations regarding the motivational component of psychological deficit in schizophrenia were suggested.

REFERENCES

1. BORING, E. G. Learning in dementia praecox. *Psychol. Monogr.*, 1913, **15**, No. 2 (Whole No. 63).
2. CAMERON, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 861–921.
3. HART, B., & SPEARMAN, C. Mental tests of dementia. *J. abnorm. Psychol.*, 1914, **9**, 217–264.
4. HUNT, J. McV., & COFER, C. N. Psychological deficit. In J. McV. Hunt, (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 971–1032.
5. HUSTON, P. E., & SHAKOW, D. Learning in schizophrenia: I. Pursuit learning. *J. Pers.*, 1948, **17**, 52–74.
6. HUSTON, P. E., & SHAKOW, D. Learning capacity in schizophrenia. *Amer. J. Psychiat.*, 1949, **105**, 881–888.
7. KRAEPELIN. E. *Dementia praecox and paraphrenia*. R. M. Barclay (Trans.). Edinburgh: Livingstone, 1919.
8. KENDIG, ISABELLE, & RICHMOND, WINIFRED V. *Psychological studies in dementia praecox*. Ann Arbor, Mich.: Edwards, 1940.
9. MALAMUD, W., & PALMER, ELEANOR M. Intellectual deterioration in the psychoses. *AMA Arch. Neurol. Psychiat.*, 1938, **39**, 68–81.
10. PAINTAL, A. S. A comparison of the galvanic skin responses of normals and psychotics. *J. exp. Psychol.*, 1951, **41**, 425–428.
11. PETERS, H. N. Multiple choice learning in the chronic schizophrenic. *J. clin. Psychol.*, 1953, **9**, 328–333.
12. ROE, ANNE, & SHAKOW, D. Intelligence in mental disorders. *Ann. N. Y. Acad. Sci.*, 1942, **42**, 361–490.
13. SANDS, S. L., & RODNICK, E. H. Concept and experimental design in the study of stress and personality. *Amer. J. Psychiat.*, 1950, **106**, 673–679.
14. SHAKOW, D. The nature of deterioration in schizophrenic conditions. *Nerv. ment. Dis. Monogr.*, 1946, No. 70.
15. WITTMAN, M. PHYLLIS. An evaluation of opposed theories concerning the etiology of so-called "dementia" in dementia praecox. *Amer. J. Psychiat.*, 1937, **93**, 1363–1377.

Received March 7, 1955.

TIME AND SPACE ORIENTATION IN SCHIZOPHRENICS

C. O. DE LA GARZA AND PHILIP WORCHEL

University of Texas

FEW systematic investigations (3, 5, 6) have been conducted to validate the conclusions often expressed in the literature that the psychotic patient frequently loses orientation to time and place. Clinical observations have often been used to justify such a contention, and routine psychiatric examination (2) almost always includes such simple questions as, "What day (month and year) is it?" and "What place is this?" On the basis of these limited questions, generalizations are drawn as to the intactness of the orientational ability and the present extent of a psychotic or organic condition. It thus becomes of major importance, not only for practical diagnostic considerations but for a better understanding of the nature of the psychotic process, to explore systematically and under controlled conditions the degree and type of loss in orientational skills in psychotic patients.

PROCEDURE

Subjects

Two groups of fifty *S*s each were selected from the Austin State Hospital: an experimental group consisting of patients diagnosed as paranoid schizophrenic, and a control group made up of attendants and applicants for the position of attendant.

The patients were selected on the basis of the following criteria: (*a*) residence at the hospital for at least two years but not more than ten, (*b*) age not over 55 years at the time of the test, (*c*) freedom from organic concomitants, and (*d*) no electroshock, insulin coma, or similar somatic treatment within the past two weeks.

Each of the 50 control *S*s was matched as closely as possible with each subject of the experimental group on Wechsler-Bellevue Vocabulary score, sex, and age. There were 25 males and 25 females in each group.

Since Shakow (8) has suggested that many of the "diagnostic" differences in the test performance of schizophrenics may possibly be due to lack of cooperation, each subject was rated by the senior author from one to six on Shakow's (8) scale of cooperativeness as modified by Moran (7). Of the 50 control and 50 schizophrenic *S*s, 45 and 38 respectively received ratings of 4 and above ("performed with more than moderate interest"). A chi-square test of the ratings on cooperativeness yields a value of 3.48, which does not reach the .05 level of confidence and suggests that the two groups may be considered as equated for cooperativeness.

Time Orientation Tests

The first time-orientation test consisted of 20 questions on general time concepts (GTC). The subject was asked to answer the following questions: 1. Is it morning or afternoon? 2. What day of the week is it? 3. What month is it? 4. What year is it? 5. About what time is it by the clock now? (Correct if within 30 minutes) 6. What day of the month is it? 7. What season of the year is it? 8. How many minutes are there in an hour? 9. How many months are there in a year? 10. How many hours are there in a day? 11. How many seasons in a year? 12. How many seconds in a minute? 13. In what month is Christmas? 14. What does one mean when he says "9 A.M." and "9 P.M."? 15. In what month is Thanksgiving Day? 16. What day follows Thursday? 17. How many days are there in a week? 18. How many days are there in a year? 19. Today is ———. What day is it in New York? 20. What month follows September?

The second time-orientation test consisted of ten sets of items. The three components of each set were related to each other in that they were part of a common group of events, such as three days in a week, or three well-known events or concepts that occur in a definite temporal order. The items were typed on 3 x 5 cards, and each set was presented to each *S* in a mixed order. The order of presentation was the same for all *S*s. The *S* was asked to rearrange the following sets so that the items or events were in their proper order:

1. May March April 2. Sept. 17, 1951 April 10, 1950 December 29, 1952 3. Winter Autumn Summer 4. Tuesday Saturday Friday 5. Declaration of Independence Pilgrims Settle in America Columbus Discovers America 6. New Year Christmas Easter 7. October July February 8. Today Yesterday Tomorrow 9. Future Past Present 10. Washington Roosevelt Lincoln

Space Orientation Tests

Three space-orientation tests were used: bodily orientation, perceptual orientation, and geographic orientation.

Bodily orientation refers to the conscious localization of stimuli within or on the body. The following 10 questions were presented to *S*: 1. Which is the back of your body? 2. Which way is down? 3. Which is your left foot? 4. Where have I placed the wood block (right shoulder)? 5. Which way is up? 6. Which is the front of your body? 7. Which is your right foot? 8. Which is your left hand? 9. Where have I placed the wood block (left knee)? 10. Which is your right hand?

Perceptual orientation refers to the location of objects usually situated in the immediate perceptual field on the basis of a right-left, up-down, front-behind frame of reference. The *S* was required to rearrange two or three of the Binet blocks, which were set before him on a flat surface in either a perpendicular or horizontal arrangement. The blocks were set in a straight line, approximately 2 in. apart. The ten tasks were: 1. Give me the block nearest to you. 2. Put the block farthest from you in front of the block nearest to you. 3. Give me the block in back. There is one in front and one in back. Give me the one in back. 4. Give

me the block on your right. 5. Put the block nearest to you behind the block farthest from you. 6. Put the block in back under the block in front. 7. Give me the block farthest from you. 8. Give me the block in front. 9. Put the block in front on top of the block in back. 10. Give me the block on your left.

Geographic orientation refers to the location of objects with respect to compass positions. While this system is usually reserved for objects outside the immediate perceptual field, it may also be used for specifying the positions of objects within the immediate perceptual situation. A series of twenty-two subtests modifying those used by Lord (4) was employed.

In subtest A, the *S* was asked to indicate by pointing: 1. north, 2. east, 3. south, 4. west, 5. northeast, and 6. southwest. A reply was considered correct if *S* indicated a direction within a 90° sector (±45°).

In subtest B, the subject was asked to locate some cities or towns relative either to his place of longest residence or to Austin, Texas. At least five locations were obtained, and if directions given fell within a 90° quadrant of the correct point, the response was considered correct.

In subtest C, the *S* was asked to tell the examiner in which compass direction he would have to go to get from the *S*'s last longest place of residence to another city. Five cities were given: 1. Austin (if Austin were given as home, El Paso was substituted), 2. Fort Worth, 3. New Orleans, 4. Washington, D. C., and 5. Berlin, Germany.

In subtest D, the *S*s were required to indicate the compass direction between two given points. The instructions were:

> What direction would you have to go to get from 1. St. Louis, Missouri, to New York City? 2. Chicago, Illinois, to Nome, Alaska? 3. New York City to Paris, France? 4. Seattle, Washington, to New Orleans, Louisiana? 5. Denver, Colorado, to San Francisco, California? 6. Rome, Italy, to Madrid, Spain?

Before the time-orientation and space-orientation tests were given, the following information of a general nature from each *S* was obtained: name, place of birth, age, education, usual occupation, residence prior to coming to Austin, and place of longest residence.

Results and Discussion

The mean and the *SD*'s on all the orientation tests for both groups are presented in Table 1. Since almost all the distributions for the control subjects were highly skewed and truncated, the sign test was applied to the paired scores in order to test the hypothesis that the median of the population of differences is zero. Since there were 50 paired scores for each time- and space-orientation test, the number of differences which were positive and negative was calculated, and a χ^2 test with one degree of freedom was applied (corrected for discontinuity). Ties were omitted. Table 2 shows the number of ties and positive and negative differences on each time and space orientation test together with the chi-square values. For a two-sided test, the chi-square values for the time-orientation tests I and II and space-orientation test II were significant beyond the .01 level. For space-orientation test III, the chi-square value of 6.90 was significant at the .02 level. For space-orientation test I, the χ^2 value was not calculated since the theoretical frequency for positive and negative differences was only 2.0. Fisher's exact probability method was employed on the dichotomized bodily orientation (I) data (all right vs. one or more wrong). All the normals scored perfectly, while only four schizophrenics missed one or two items. The probability of obtaining these proportions for the two groups is not significant at the .05 level. Thus, the ability to localize stimuli on the body and to designate accurately the various parts of the body are apparently not significantly impaired in the schizophrenic. The fact that this test did not discriminate among the normals suggests perhaps that the items were too simple, and that impairment, if present, in bodily orientation would therefore

TABLE 1

Performance on Orientation Tests by Schizophrenic and Normal *S*s

Test	No. Items	Schizophrenic		Control	
		M	*SD*	M	*SD*
Time orientation—I & II	30	20.95	1.64	24.08	3.28
I Gen. time concepts	20	15.24	1.31	19.14	1.12
II Groups of related events	10	5.77	2.56	9.14	1.38
Space orientation—I, II, III	42	27.95	1.92	35.50	5.03
I Bodily orientation	10	9.83	0.62	10.00	0.00
II Perceptive orientation	10	7.48	1.94	9.54	0.96
III Geographic orientation	22	10.64	1.51	15.42	1.53

TABLE 2

Number of Differences Between Paired Control and Schizophrenic Subjects on Each Orientation Test

Direction of Difference	Time		Space		
	I	II	I	II	III
Positive	44	47	4	32	33
Negative	2	0	0	4	14
Ties	4	2	46	14	3
χ^2	36.54	45.02	—	20.26	6.90

not show up in the schizophrenics. It is also possible that orientation involving a concrete, specific frame of reference would be less amenable to deterioration. Support for the latter hypothesis is found in the analysis of items in the other space and time orientation tests.

More specific information about the nature of the differences between the groups than that provided through the comparison of the tests was obtained through an item analysis. The responses of all subjects were examined to determine which test items were failed most frequently by each group.

The relative power of each item for discriminating the members of the control group from those of the schizophrenic group was obtained by the χ^2 method. Thirty-nine of the 72 original test items obtained p-values beyond the .05 level. These were the items to which schizophrenic subjects were typically unable to reply correctly. It is not completely clear why these 39 items should be more difficult for the schizophrenic subjects. On the first time-orientation subtest (GCT), 12 of the 20 items in the subtest were more difficult for the schizophrenic subjects. The eight items which were "easy" for the schizophrenics seem to be those that are overlearned or have some concrete referent. Thus, "How many minutes in an hour?" or "How many months in a year?" or "What day follows Thursday?" are items of information repeated frequently throughout life, and no new learning is required. The first item, "Is it morning or afternoon?" is answered correctly almost always. The schizophrenic can observe directly whether it is morning or afternoon. When learning is required, or a referent is needed that is not in the immediate perceptual field, the schizophrenic encounters difficulty. For example, "What day of the week is it?" is significantly more difficult for the schizophrenic.

It may be, however, that such temporal items are difficult because of institutionalization. The hospital routine of each day is so much like that of any other that there are no unique events to distinguish the days or sufficient motivation to remain oriented in this sense. That this is not a completely valid explanation is attested to by the fact that it has been demonstrated clinically that schizophrenic patients are frequently disoriented on arrival at the hospital and become oriented prior to discharge. Of course, there is no way of proving that these particular patients were disoriented at the time of admission. One way to settle this issue is to use a control group of institutionalized nonpsychotic subjects such as tuberculosis patients or prisoners.

It is also possible that the "loss" in orientational skills of the schizophrenic subjects may be due to their inability to comply with instructions because of such psychotic factors as autistic preoccupation or irrelevant affective processes. The present study was not designed to control such factors, but they must be taken into consideration and eventually investigated.

On time-orientation test II, groups of related events, both groups did equally well on only one item (May, March, April). Of the ten sets of cards which make up this subtest, the one just mentioned is the only one which presents a series of temporally related events that is neither complicated in any way nor has omitted some intermediate portion of the series. The series "April 10, 1950," "September 17, 1951," "December 29, 1952," however, is made complicated by having three elements to be considered simultaneously: months, days, and years. The series "Tuesday," "Friday," "Saturday" is complicated by the omission of "Wednesday" and "Thursday." A series of temporal items that are successively related, like May, March, April, contains the referent cues within the series; for example, March is associated successively with April, and April with May. When some frame of reference is required, however, that is not given in the immediate perceptual field, that is, contained within the series itself, the schizophrenic seems to find it difficult to orient the elements properly.

Perceptual orientation, the second of the space orientation subtests, was composed of 10 items. Both groups did equally well on five items. These five items were all similar in that the patient was asked to give the examiner a block located in terms of its relation to *S*. On the other five, which were failed more often by the schizophrenics, not only were there two frames of reference to be considered, the *S* and a particular block, but the instructions were complicated by requiring *S* to retain three locational cues, as for example, "Put the block *farthest* away from you in *front* of the block *nearest* to you." Here the failure by the schizophrenics may simply reflect an impaired attention span or an inability to retain complexly related items.

Inability to perceive the relationship apart from attention span may be tested by giving the patient written instructions or by frequent verbal repetition of the instructions. One aspect of deterioration may be an impaired attention span in addition to a decrement in perceiving and comprehending complex temporal and spatial relationships.

The last of the space-orientation subtests, geographic orientation, was made up of 22 questions. On 13 of these items, the schizophrenic *S*s did significantly poorer, a finding which supports the hypothesis of the difficulty in orientation among schizophrenics in the absence of a frame of reference in the immediate perceptual field.

The final six items in this group asked the subject to give the direction between two geographic locations, neither of the two locations being the city where the subject was then being tested. On only one item did the control subjects do better than the schizophrenics. The information required for the correct answers to these questions is not of the kind that is generally available to everyone. This point is reflected in the high number of failures on these items by the control subjects. The failures by the pathological groups therefore cannot be evaluated in terms of possible loss due to their mental illness.

The present findings of significant differences between the control group and the schizophrenics confirm Minkowski's clinical observations of disorientation in these kinds of patients (6). Whereas Fisher's patients were observed to be disoriented while in acute stages of schizophrenia (1), the present findings suggest that the loss in orientational ability suffered by the schizophrenic in the acute phase may persist as a characteristic of the schizophrenic syndrome.

Israeli (3) concluded from his studies of 37 paranoid schizophrenics that the notion of time rarely disappeared. On the other hand, the schizophrenic subjects in the present study differed significantly from their controls. Since Israeli did not include a control group in his study, his conclusions represent subjective evaluations of his data. In addition, negative results might be due to the nature of the items employed. For example, 8 of the 20 items in the first time orientation test of the present study were passed by both the schizophrenic and normal subjects.

Summary and Conclusions

The purpose of this study was to compare the spatial and temporal orientation of a group of 50 schizophrenic subjects with a matched group of 50 normal subjects selected from attendants and applicants for the position of attendant at the Austin State Hospital. The groups were matched in terms of Wechsler-Bellevue Vocabulary raw scores, age, sex, and economic status. The age-range of the subjects was 18 to 55 years. The experimental subjects had been hospitalized for a minimum of two years and for not more than ten years. Ratings on cooperativeness were obtained on all *S*s. The results show that:

1. The schizophrenic *S*s were significantly poorer than the normals on all the space- and time-orientation tests except that for bodily orientation

2. Thirty-nine of the 72 items differentiated the schizophrenic from the control *S*s.

3. These 39 "failed" items seemed to have in common the absence of a frame of reference in the immediate perceptual field or to require attention to and retention of complex relationships.

It was suggested that the defects in orientation in psychotics may be subsumed under the general notions of loss in conceptualization and lowered attention span involving complex relationships.

REFERENCES

1. Fisher, Franz. Zeitstruktur und schizophrenie. *Zeit. f. d. Neurol.*, 1929, **121**, 544–574.
2. Henderson, D., & Gillespie, R. D. *A text-book of psychiatry* (7th Ed.). London: Oxford Univer. Press, 1950.
3. Israeli, N. *Abnormal personality and time.* New York: Israeli, 1936.
4. Lord, F. E. A study of spatial orientation of children. *J. educ. Res.*, 1941, **34**, 481–505.
5. Minkowski, E. *La schizophrenie.* Paris: Alcon, 1927.
6. Minkowski, E. La notion du temps en psychopathologie. *Evolution Psychiat.*, 1929, Ser. II, No. 1, 65–85.
7. Moran, L. J., Jr. Vocabulary knowledge and usage among normal and schizophrenic subjects. *Psychol. Monogr.*, 1953, **67**, No. 20 (Whole No. 370).
8. Shakow, D. Some psychological features of schizophrenia conditions. In Reymert, M. L. (Ed.) *Feelings and emotions. The Mooseheart Symposium.* New York: McGraw-Hill, 1950.

Received March 14, 1955.

INTRA-INDIVIDUAL VARIABILITY IN SENTENCE COMPLETION RESPONSES[1, 2]

JERRY OSTERWEIL[3]
The Menninger Foundation
AND DONALD W. FISKE
The University of Chicago

INTRA-INDIVIDUAL variability in responses refers to the differences between the responses of one person to the same stimulus presented at different times, differences above and beyond systematic trends over time. The present study, a part of a larger research program (5, 7), involved response variability on a sentence completion test. In addition to investigating the nature and frequency of changes on a projective test, where responses are free rather than multiple choice, this study had two specific objectives: (*a*) to determine whether variability was in part a function of the sentence stem; (*b*) to determine whether variability was associated with the degree of individuality in a subject's responses.

The concept of intra-individual variability has appeared like an unwelcome but persistent intruder in psychological writings. A recent review of the literature by Fiske and Rice (5) has demonstrated how this concept has recurred in many areas of theory and research. But the general attitude has been to reduce or minimize its significance or effect. Until recently, it has rarely been the major emphasis in an empirical study.

Historically, part of the neglect of the concept of intra-individual response variability grew out of a tradition in psychometric theory to regard the variance of a test as consisting of two components, the variance of the true score and the error variance. While there can be no doubt that random errors of measurement reduce the reliability of a test, several writers [such as Spearman (10), Jackson and Ferguson (6), Brown and Thomson (1), and Burt and Stephenson (2)] have suggested that error variance may also include a systematic component based on individual differences in variability. The empirical study which bears most directly on this problem, however, is an unpublished doctoral dissertation by Harold Fagin (4). This study analyzed the retest unreliability of two batteries of interest and trait scales into (*a*) a random error variance and (*b*) a consistent behavioral or personal variance manifested in the form of reliable individual differences.

If there are such consistent individual differences in variability, then a number of important questions can be raised. Are consistent differences in variability of response manifested on other tests, such as the sentence completion test? Is response variability a function not only of the person but also of the stimulus? What psychological variables are associated with differences in intra-individual response variability? This study attempts to throw light on some of these problems.

PROBLEM AND PROCEDURE

The major question in this study was whether responses which have a high frequency of occurrence in a group and are therefore based on a greater degree of social consensus are changed less than responses occurring infrequently. The expectation was that popular responses (the single set of equivalent responses with the highest frequency of occurrence to any given sentence stem) would be less variable than unique responses (responses given by only a single member of a group).

A popular response is conceived as a response which is determined largely by the characteristics of the stimul[illegible] its more stereotyped, socially shared meanings i[illegible]ne culture. A unique response is likely to be base[illegible] on a more individualistic, subjective source. Since the stimulus does not change and the stereotyped, socially shared meanings do not change, but the state of the individual can and does change, responses that are more subjectively determined (such as unique responses) should change more than more objectively determined responses (such as the populars).

Subjects and Testing Procedure

The same form of sentence completion test was given under anonymous conditions on two occasions to four

[1] This research was supported in part by the United States Air Force under Contract Number AF 18(600)-601, monitored by the Director of Research, Personnel Research Laboratory, Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

[2] We are grateful to Drs. Shib Mitra and Laura Rice for serving as judges in this experiment.

[3] This paper is based in part upon the dissertation of the first author, who acknowledges with thanks the assistance of Dr. Howard Hunt and Dr. Hedda Bolgar.

groups. The numbers of subjects were 21, 14, 38, and 20 respectively (93 in all).

All four groups consisted of students enrolled in some kind of basic psychology course. Group 1 was clearly the most homogeneous group since all the subjects were white female student nurses in late adolescence or early adulthood. The administrations for Group 1 were two weeks apart, whereas the other groups had a three-week interval between testings. There is no reason to believe that this difference in the interval had any systematic effects on the responses. The other three groups were composed largely of white males and females in late adolescence or early adulthood. These groups did not differ significantly in age. There were a few individuals in later adulthood and a few Negroes in each of these three groups. These three groups were composed of extension students enrolled in psychology courses at a state university. Analyses showed no significant sex differences in average change per person and per sentence stem.

Groups 1 and 2 took the test under standard instructions which suggested that the study concerned group attitudes. Groups 3 and 4 received instructions to assume a specific set in answering the test. Since the major findings of this study held for these last two groups as well as for the groups with standard instructions, the special instructions will not be discussed here.[4]

While each group received the same set of instructions on each occasion, all groups were told before the second testing that they should give the completions that occurred to them at that time. To sustain motivation they were also informed, on the second testing, that the research was sponsored by the armed forces.

Rating Procedure

Relative response frequencies were determined for each group separately. For a particular sentence stem, all responses from one group of subjects were first classified by one judge into sets of similar or equivalent completions. From this classification it was possible to determine empirically the frequency of each set of responses in the group.

Similar completions were defined as completions with the same key idea and no more than slight differences in any adjectives or phrases that qualified the main idea. From this initial classification made by the first judge, a sample of responses to 15 sentences (for 38 individuals) was drawn, the sample being stratified on the basis of relative frequencies of response. To determine the reliability of judgments, two additional judges classified these responses in the stratified sample. The judgments of each of the two judges, who were given appropriate training first, were compared with the initial judge's classifications, which were used as a criterion measure because of his superior experience with the rating procedures. He alone classified all the responses.

It was found that the judges used the same categories with very few exceptions. An analysis was made to determine the degree of agreement between judges on the frequencies of popular response and of unique responses per subject and per sentence. The scores derived from he classifications of the second and third judges (each taken separately) correlated with the scores of the first judge .74 to .98.

Three major classes of responses based on relative frequencies were employed in this study:

1. Popular responses were the single set of responses to a given sentence stem which had the highest frequency in a given group. If for a given sentence two or more sets of responses were tied for highest frequency, a table of random numbers was used to select the popular. All other sets of responses with frequencies greater than one would then become the common responses.

2. The common responses were those sets of responses that were given by more than one individual in a group, excluding the popular.

3. Unique responses were those sets of responses which had frequencies of one, i.e., which were only given by one member of a group.

A similar procedure was used to determine the interjudge reliability of change scores. For each pair of completions, the second response was classified as either equivalent to or different from the first one. It was considered equivalent if it repeated the first response more or less verbatim or involved only a slight change in literal meaning.

One judge rated all pairs of completions. From his ratings, distributions of frequencies of change per person and per item were constructed. Then a sample of nine subjects (50 completions each), stratified on total change score, was drawn.

After an initial training period, the two additional judges classified each pair of completions in the sample. The frequencies of changed responses per subject and the frequencies per sentence stem for each of these judges were compared with those of the first judge. The interjudge reliability of changed responses for both subjects and sentences was again consistently high. (Product-moment correlations ranged from .80 to .96.) It may be noted that about 80 per cent of the responses were changed on the second occasion. Thus change, rather than stability, is characteristic of sentence completion responses.

RESULTS

It was first necessary to determine whether the response classes employed were internally consistent. Reliabilities were therefore computed for the number of popular, unique, and changed responses per individual in each of the four groups, using Kuder-Richardson Formula 20 (11, p. 92). With the exception of one coefficient of .90 and one of .06, the coefficients ranged from .43 to .72, with no systematic differences among groups or among classes of responses. This finding supports the underlying assumption of this research that people show consistent individual differences in their variability and in their tendency to give popular and unique responses. Not all the residual variance can be assigned to error, therefore; some of it has a systematic source in the in-

[4] For a full exposition of these and other details, see (8).

dividual. These findings provide the basic justification for undertaking this research.

For each pair of groups, product-moment correlations were computed between the frequencies of changed responses per sentence. The median of the six correlations was .48. Similar sets of intercorrelations were computed for both popular and unique responses. The median intercorrelations were .50 and .38 respectively.

All but one of these 18 intercorrelations was statistically significant. Thus, there are consistent differences from stimulus to stimulus in the frequencies of popular, unique, and changed responses evoked.

The relationship between change measures and measures of relative frequency of response was determined. These correlations can be computed over persons and also over stimuli. Table 1 shows the correlations over persons for each group. The three scores are the frequencies of unique, popular, and changed responses per person. In spite of sampling fluctuations, the directions of the correlations are quite consistent from group to group: persons with high change scores tend to have high unique scores and low popular scores.

Parallel correlations were computed over sentences (see Table 2). It will be seen that sentences with a high number of changed responses tended to have relatively high frequencies of unique responses and relatively low frequencies of popular responses.

These analyses show that the number of changed responses is related to the frequencies of unique and popular responses. It remains to be seen whether, on the average, the probability of a unique response being changed is greater than that of a popular response. To answer this question, the proportion of unique responses that were changed was compared to the proportion of popular responses that were changed for each sentence and for each group. Thus, for Group 3, there were 33 sentences for which the proportion of unique responses changed was higher than the proportion of populars changed. Similar comparisons, using the Sign Test (3, pp. 247–251, 324), were made involving the common responses, and the combined popular plus common classes (see Table 3).

Analyzed sentence by sentence, proportionally more unique responses than popular responses are changed. The direction of the difference is the same for all four groups. This result, however, fails to reach statistical significance for Groups 2 and 4.

Also analyzed sentence by sentence, proportionally more unique responses changed than responses in the popular-plus-common combination. This trend appeared in three groups, but the result was not statistically significant in Group 1. Thus responses given by only one

TABLE 1

PEARSON CORRELATIONS OF CHANGE SCORES FOR PERSONS WITH POPULAR SCORES AND UNIQUE SCORES

Change Scores vs.	Groups			
	1 (N = 21)	2 (N = 14)	3 (N = 38)	4 (N = 20)
Popular score	−.41*	−.21	−.72**	−.47*
Unique score	+.51**	+.51*	+.24	+.35

* Significant at the .05 level.
** Significant at the .01 level.
Note.—One-tailed tests of significance were used throughout this study.

TABLE 2

PEARSON CORRELATIONS OF CHANGED RESPONSES PER SENTENCE WITH POPULAR RESPONSES AND WITH UNIQUE RESPONSES
(N = 50)

Changed Responses vs.	Groups			
	1	2	3	4
Popular responses	−.43**	−.26*	−.43**	−.14
Unique responses	+.53***	+.39**	+.80***	+.77***

* Significant at the .05 level.
** Significant at the .01 level.
*** Significant at the .001 level.

TABLE 3

SIGNIFICANCE OF DIFFERENCES BETWEEN PROPORTIONS OF CHANGED RESPONSES FOR VARIOUS PAIRS OF RESPONSE CLASSES
(Entries are p-values for One-tailed Sign Test)

Predicted Direction of Differences Between Classes	1	2	3	4
Unique > popular	.025	N.S.*	.01	N.S.
Unique > popular plus common	.125	.025	.01	.05
Unique > common	N.S.	N.S.	.025	N.S.
Common > popular	N.S.	N.S.	N.S.	N.S.

* N.S. indicates $p > .125$.

person in a group were changed more frequently than responses given by more than one person. Such trends were not significant when the unique class was compared with the common, and the common with the popular, perhaps because the identification of truly unique and popular responses is uncertain with small samples.

At first thought, one might think that these findings are obvious on the basis of simple probability theory. Suppose the relative frequency of occurrence in a group is .4 for one response and .1 for another. It might seem that the probabilities of repetition would be .16 and .01, respectively, the more frequent response being repeated more often. This approach assumes that the probability of a given response is the same for each member of a group. Analysis indicated that this assumption was clearly untenable for these data and that the proportions of repeated responses were systematically higher than those predicted by such probability considerations.

DISCUSSION

From the general formulation sketched briefly in another paper (7), certain relationships with variability scores would be predicted for such a free-response situation as a sentence completion test.

First, for a distribution of responses made by many persons to one item on two or more occasions, the larger the variation of that distribution, the more likely it is that the average person's responses will change from one trial to another. This prediction is confirmed in this study. The fewer popular and common responses (and therefore the more unique responses) a group makes to an item, the more changed responses are given to that item.

Second, in this study, consistent individual differences in popular scores for persons were demonstrated. A popular score, therefore, has internal consistency; it presumably also has psychological meaning. Thus, our rationale leads to the prediction that the higher the person's popular score, the lower his change score. This prediction was also confirmed.

While this paper has demonstrated that variability scores for a free response test can have internal consistency and can be related to relative response frequency, the trends are not as strong and the correlations are not as high as those for multiple-choice instruments (7). A number of factors account for the differences. On the sentence completion test, the responses could be classified reliably, but not with the unequivocal certainty of tests with fixed alternatives. Further, the relative frequency distributions for items were based on fairly small groups. Finally, the much higher average proportion of changed responses on the sentence completion test (around 80 per cent) probably reflects the fact that the task is relatively unstructured, and the subjects' sets are presumably more unstable, even from item to item.

In two respects, this study and the one on multiple-choice instruments (7) are similar. Both measured change in terms of differences between two trials rather than as variability over a series of trials, which would yield a more reliable measure of change. Both showed some systematic trends over time which contribute to the change scores. In the other experiment, the systematic trend over time contributed only a small part of the total variability. While the contribution of such a trend was not assessed in this study, it might well be larger for free-response tests than for multiple-choice tests. Such a trend obviously contaminates the pure variability score.

With a fairly strict criterion for change, a great deal of change from test to retest on a sentence completion test appeared. Also the tendency for a response to change was associated with the relative frequency of that response in a group of individuals.

This relationship between frequency of responses and variability over occasions can be interpreted in terms of Rapaport's discussion of the significance of the individual's general test-taking attitude and the differences between idiosyncratic and popular responses.

Regarding the Word Association Test, Rapaport (9, pp. 17–18) suggests that the more usual associations are derived from an attitude of conforming with the test instructions, whereas more idiosyncratic responses are largely determined by instinctual forces.

According to this rationale, individuals in test situations tend to produce some responses which are governed mostly by social convention and logic and which have a high degree of inter-individual consistency. From one point of view, the purest form of this kind of response is the popular response. On the other hand, some responses are more the result of

the operation of drives and affects and show less inter-individual consistency.

Such unique responses are weighted heavily in deriving clinical inferences from sentence completion protocols. Therefore there is a potential dilemma: the responses which are of most significance to the diagnostician are the least reliable, the least stable over time.

In must be stressed that change as measured in this study refers to change in manifest or literal meaning. It was not assumed that change in literal meaning necessarily involves change in dynamic interpretations: two successive responses to a sentence stem might have different content but the same clinical significance. Hence, two successive protocols might produce parallel sets of inferences and identical personality pictures. Alternatively, even though the responses are changed on the second occasion, the two protocols might conceivably lead to the same personality picture: each dynamic factor would be present in both protocols but would not necessarily be associated with the same items. It seems more likely, however, that the protocol for the second occasion would yield a personality picture which is comparable but certainly not identical with that derived from the first protocol. These pictures might supplement each other. In view of the large amount of change found in this study, and the suggestion that some adaptive change occurs, this whole problem deserves thorough empirical investigation.

SUMMARY

Variability in responses to a sentence completion test over a two- to three-week interval was examined in this study. The subjects were one group of student nurses and three groups of college students (N = 14 to 38). The relationship between change in responses and relative response frequency in a group was studied. The major results were these:

1. The content of the great majority of responses was changed to a greater or lesser degree on retest.

2. Individual differences in number of responses changed from test to retest showed reasonable reliability.

3. Differences between sentence stems in the proportion of changed responses were consistent from group to group.

4. For each sentence stem, the responses which are given by only one person in the group were changed more frequently than responses given by more than one person.

5. Consistent positive correlations were obtained between the number of unique responses made by a subject and the number of responses he changed. Similarly, negative correlations were found between the number of populars and the number of changed responses per subject.

It was concluded that change of responses on the sentence completion test can be attributed in part to factors associated with individuals and with items.

REFERENCES

1. BROWN, W., & THOMSON, G. H. *The essentials of mental measurement.* Cambridge: Cambridge Univer. Press, 1940.
2. BURT, C., & STEPHENSON, W. Alternative views on correlations between persons. *Psychometrika,* 1939, **4,** 269–281.
3. DIXON, W. J., & MASSEY, F. J., JR. *Introduction to statistical analysis.* New York: McGraw-Hill, 1951.
4. FAGIN, H. T. An investigation of the nature and magnitude of the components of retest unreliability. Unpublished doctor's dissertation, New York Univer., 1950.
5. FISKE, D. W., & RICE, LAURA. Intra-individual response variability. *Psychol. Bull.,* 1955, **52,** 217–250.
6. JACKSON, R. W. B., & FERGUSON, G. A. Studies on the reliability of tests. *Univer. of Toronto Dept. of Educ. Research Bull.* 1941, No. 12.
7. MITRA, S. K., & FISKE, D. W. Intra-individual variability in relation to test score and item. *Educ. psychol. Measmt,* in press.
8. OSTERWEIL, JERRY. The relationship of response frequency and instructional set to variability of response. Unpublished doctor's dissertation, Univer. of Chicago, 1954. (Microfilm No. T2570; $1.45.)
9. RAPAPORT, D. *Diagnostic psychological testing.* Vol. II. Chicago: Year Book Publishers, 1946.
10. SPEARMAN, C. Correlations calculated from faulty data. *Brit. J. Psychol.* 1910, **3,** 271–295.
11. THORNDIKE, R. L. *Personnel selection: test and measurement techniques.* New York: Wiley, 1949.

Received March 15, 1955.

INTERTRIAL EFFECTS OF IMMEDIATE SELF-COMMITTAL IN A CONTINUOUS SOCIAL INFLUENCE SITUATION

SEYMOUR FISHER, IRVIN RUBINSTEIN, AND ROBERT W. FREEMAN
Army Medical Service Graduate School

To THE extent that psychologists are concerned with the control of behavior, it is important to be able to isolate those variables which govern "resistance to change." It is already known that, in regard to human opinions and attitudes, modification is more difficult to effect when an individual is "ego involved" (15), when social supports are strong (7), when the stimuli are highly unambiguous (1, 5, 14). A number of recent studies have demonstrated that self-committal is, under certain conditions, an additional factor.

Hobart and Hovland (9) found that subjects who were required to commit themselves on a social issue prior to an attempt at influence were more resistant to influence by communication than a control group in which commitment was lacking. Deutsch and Gerard (6), repeating Asch's situation (2), similarly find fewer socially influenced errors occur in self-committal subjects.

The major aim of the present investigation was to examine the effects of self-committal in a *continuous interaction situation*, where the subject continues to commit himself on a number of successive, discrete issues or judgments, each common to a broad dimension. This type of approach is rarely encountered in the literature. The typical experimental study of influence analyzes the effects of a variable, such as self-committal (9), by taking one issue (e.g., the desirability of lowering the voting age) and inquiring whether one influence attempt will significantly differentiate an experimental group from a control group.

In our present investigation the objective was to study the behavior of individuals who were exposed to repeated influence attempts; some subjects were required to commit themselves immediately prior to each influence attempt, while this treatment was omitted for other subjects.

Our results permit us to present data of four sorts: (*a*) The effects of immediate self-committal for the *first interaction* or influence attempt; (*b*) the effects of self-committal at two different "distances," where distance is defined as the discrepancy between the subject's expected judgment and the influence source; (*c*) an analysis of changes occurring between interactions; (*d*) the accumulative effects of continued self-committal. It will also be of interest to follow out the meaning and implications of certain quantitative measures of social influence in the light of our results.

METHOD

The Situation

A perceptual task, involving the judgment of dot numerosity on tachistoscopically presented slides, was employed. Five slides were drawn up, each consisting of essentially the same pattern, but differing in the actual number of dots. These five slides, hereafter referred to as A through E, contained 61, 71, 98, 110, and 130 dots, respectively. The slides were arranged in an arbitrary and standard sequence of 50 exposures so that in each block of 10 exposures each slide would appear twice; an additional restriction was that the same slide never appeared twice in succession.

The experimental room contained a table and two chairs, facing a screen located against one wall of the room. The slide projector was approximately 19 feet away from the screen. All windows in the room were blacked out. Beside the slide projector were four large slide boxes, each filled with slides; these boxes were in full sight of the subjects (*S*s) and served to create the impression that each of the 50 exposures represented a new slide. That the slides were perceived as discrete is indicated by the fact that less than one per cent of our *S*s showed any evidence of being aware of slide repetition.

Subjects

The *S*s were chosen from enlisted men who were about to be separated from the armed forces at Fort Meade, Maryland. Because of administrative difficulties in obtaining men, our sample must be considered as "accidental" (13), and no claims to randomness can be made. Furthermore, such variables as rank, education, age, and intelligence remained relatively uncontrolled.

Procedure

Four groups of *S*s were used in the study.

Control Group. Fifty-one *S*s were requested to make estimates of the number of dots on each of the 50 exposures. Each *S* was tested individually without any suggestions or influence attempts. The *S* was brought into the room by the experimenter (*E*) and was seated at a table facing the screen. He was then told that he would be shown 50 different slides, each having a number of dots on it. His task was to call out the correct number of dots on each slide. The slides were exposed for 3 seconds apiece.

After the *Ss* were tested, the data were analyzed to obtain means and standard deviations for each of the 50 exposures. Thus, for each exposure we had available an estimate of the expected group mean and the degree of group variability. These results served as normative data, indicating group performance in the absence of any influence attempts.

Noncommittal Group. Thirty-two *Ss* were tested individually in the presence of a civilian "role-player" (hereafter referred to as *P*), introduced to *S* as a person who was also about to be tested. The *S* and the *P* were brought into the experimental room by *E*, and were seated side by side facing the screen. The *E* then read the instructions, which were designed to endow *P* with a certain amount of initial "expert prestige," by informing *S* that *P* had completed a six-month training course. The instructions concluded with:

> In order to save time, I'm going to work with the two of you together. Remember, you will see each slide for only 3 seconds. After the slide is removed, call out the number of dots on that slide. Make your judgments in units of 5—that is, 5 dots, 10 dots, 15 dots, etc.; don't try to decide the *exact* number of dots, just to the nearest 5. For example, if you think there are 12 dots, say 10; if you think there are 14 dots, call out 15. [These examples were also used in the Control Group instructions.]
>
> Let's see, since you (*P*) have already had the training, we'll have you call out your answers first, and then you (*S*). But don't pay any attention to each other's answers; each of you call out how many dots you yourself believe are on the slide. Pay close attention, because it's important for both of you to do as well as you possibly can on this. Any questions?

It will be observed that in this procedure, *P always gave his estimate before S*. The *P*'s estimate for any one exposure was the same for all *Ss* in this group, and was derived from the normative data obtained from the Control Group. For each of the first 40 exposures (Blocks 1 through 4) *P*'s estimate consisted of the Control Group mean plus one standard deviation above that mean. For the last 10 exposures (Block 5), *P*'s estimate was based on the Control Group mean plus five standard deviations above the mean.[1]

The 50 predetermined judgments which were given by *P* were previously memorized (or read from a concealed "crib sheet" held by *P*.)

Committal Group (Group Distance). Forty-four *Ss* were tested with a procedure identical to that of the Noncommittal Group except that *S* was requested to give two estimates to each exposure, one before *P* and one after *P*. However, regardless of what *S*'s first estimate was, *P* called out the same number as he had given in the noncommittal situation.

This committal group is labeled "Group Distance" (GD) to indicate that *P*'s estimate is completely independent of any one *S*'s estimate, but—assuming a sample of *Ss* comparable to the Control Group—*P*'s estimate for each exposure should represent a constant distance for the *group*.

In drawing up the instructions for our Committal Groups, there was a question of precisely how to structure the situation. Since there was no a priori evidence to use as a guide, it was finally decided to try two sets of instructions, differing only in the reason offered to *S* why two estimates were requested. Subsequent analysis showed that the two sets produced essentially the same results, and the data were pooled.

Committal Group (Individual Distance). Forty-one *Ss* were tested, using the same procedure and instructions as in the first Committal Group (GD), with *S* making pre- and post-*P* estimates to each slide. In this group of *Ss*, however, *P*'s estimates were always contingent upon *S*'s first estimate to an exposure: instead of adding standard deviation units to the Control Group mean, the standard deviation units were added to *S*'s *individual and actual estimate*.

RESULTS

Since the *Ss* in the Noncommittal and GD Committal Groups were exposed to identical estimates by *P*, the mean post-*P* judgments of the GD Committal Group can be compared with mean responses of *Ss* in the Noncommittal Group. In Table 1, these means, along with *t* tests, are presented for the 10 exposures in Block 1 (at a distance of 1 *SD* unit) and the 10 exposures in Block 5 (distance of 5 *SD* units).

It can be seen that in the first 10 trials, there are no outstanding differences between the two groups—with the exception that on the very first exposure (Slide E-1 in Block 1) the Noncommittal *Ss* tended to show more influence (i.e., higher mean response) than the Committal Group.[2] The probability levels (p)

[1] The use of standard deviations as distance units was suggested by the following considerations. In general, within-group variability offers an index to "certainty" or "confidence"; i.e., if there is no variability on a particular stimulus, this means that all *Ss* gave the same response; unless the experimental instructions or "set" forced this specific response, the lack of variability must indicate that the stimulus is extremely unambiguous, and the *Ss* quite certain of the validity of their response. Conversely, marked variability indicates stimulus ambiguity and corresponding subject uncertainty. If distance units are defined in absolute terms (e.g., the mean plus 50), "equal" distances may be quite unequal psychologically on exposures with different variances. 50 points added to the mean of an unambiguous stimulus might represent a much greater psychological distance than 50 points added to a generally ambiguous stimulus.

[2] It will be noted that, although there were 130 dots on the slide, the mean response to Slide E-1 is conspicuously low in both groups, even after influence. This is probably a function of the instructions, which used low-numbered examples, thus establishing a set to underestimate. Data from the Control Group reveal that these *Ss* similarly underestimated the true number of dots (mean judgment on first slide was 69). This effect provides an excellent illustration of the need for normative data when no knowledge is available to permit

TABLE 1

JUDGMENTS OF NONCOMMITTAL SUBJECTS COMPARED WITH POST-*P* JUDGMENTS OF GD COMMITTAL SUBJECTS

Slide		Block 1			Block 5		
		Non-comm.	Comm.	*p*	Non-comm.	Comm.	*p*
A-1	Mean	71.7	68.8	NS	109.8	93.3	.02
	SD	17.2	19.9		28.9	27.7	
A-2	Mean	68.9	65.7	NS	102.9	98.9	NS
	SD	14.4	16.1		31.2	36.0	
B-1	Mean	76.9	73.5	NS	124.1	104.9	.05*
	SD	20.4	28.5		45.6	30.4	
B-2	Mean	80.2	74.5	NS	136.2	105.1	.01*
	SD	27.9	19.9		57.3	26.0	
C-1§	Mean	163.1	157.2	NS	*255.3*	*169.5*	*.001**
	SD	27.9	34.9		109.2	30.3	
C-2	Mean	152.3	152.7	NS	287.8	223.5	.01*
	SD	21.4	26.4		101.4	66.0	
D-1	Mean	145.6	144.9	NS	255.0	214.9	.05*
	SD	31.8	30.8		85.4	51.2	
D-2	Mean	153.1	156.2	NS	245.0	228.1	NS
	SD	19.5	25.8		84.9	75.5	
E-1§	Mean	*111.1*	*97.5*	.07	301.8	264.1	NS*
	SD	16.9	43.1		116.5	44.6	
E-2	Mean	173.6	175.9	NS	325.3	293.2	NS
	SD	36.0	47.3		110.9	90.4	

§ Slide E-1 was the very first slide in the series of 50 trials; Slide C-1 appeared on the 41st trial (i.e., the first exposure in Block 5).

* Heterogeneity of variance significant.

appearing in the table are all based on two-tailed tests; one-tailed tests, which might be justified on a theoretical basis, would yield a significant difference at the .05 level on Slide E-1.

In Block 5, where the distance is greater, the effects of self-committal become more marked, as evidenced by the fact that on all 10 exposures the mean judgments of the Noncommittal Group are higher than the Committal Group, 6 of the 10 two-tailed tests being significant at better than the .05 level of confidence. Again we note that the difference between the two groups is most conspicuous on the 41st exposure (Slide C-1)—the first exposure at the new distance.

Table 1 further indicates differences in the variability between the two groups. *F* tests of the variances reveal that, in Block 5, the groups differ significantly on 6 of the 10 exposures (those marked with asterisks), the variance being moderately correlated with the mean.

inferences of expected responses; serious error would have resulted from the assumption that, in the absence of influence, the mean response would approximate the actual number of dots on the slide.

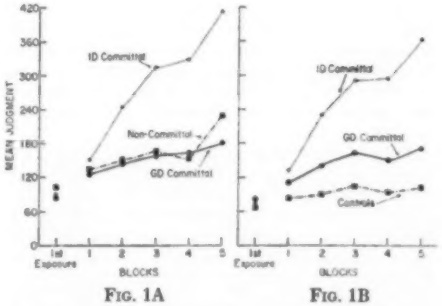

FIG. 1A FIG. 1B

FIG. 1A. RESPONSES FOLLOWING SUCCESSIVE INFLUENCE ATTEMPTS

FIG. 1B. RESPONSES PRECEDING SUCCESSIVE INFLUENCE ATTEMPTS (NO INFLUENCE FOR CONTROLS)

While it is true that the 10 exposures in either block do not permit *independent* tests of significance (since the same *S*s responded to all 10 slides), the general uniformity of the results suggests that, when the distance is relatively great, *S*s who commit themselves are less influenced than Noncommittal *S*s. At the smaller distance, however, it appears that this difference is manifested only in the initial interaction; on subsequent interactions at the same distance, both groups show essentially similar post-*P* behavior.[3] It further appears that the difference between Committal and Noncommittal *S*s is most marked when an abrupt shift is made from one distance to a larger distance (the first slide on Block 5).

The over-all picture of the effects of self-committal are best seen by ignoring differences in stimulus ambiguity (i.e., the actual number of dots on the five different slides),[4] and examining the behavior of our groups on all the blocks. In Fig. 1 the mean responses to the 10 exposures on each block have been averaged to give one over-all figure. Because of the importance of the initial interaction (on Slide E-1), this mean response is presented separately. In Fig. 1A, which compares the

[3] The mean responses shown in Table 1 for Block 1 are typical of the results in the first four blocks (all at a distance of 1): except for the initial exposure on Block 1, no differences approaching significance are found in the 40 exposures. The similarity of the two groups on the first four blocks is seen in Fig. 1A.

[4] Relationships between stimulus ambiguity and social influence will be reported in a later paper. In general, our results clearly substantiate the findings of Asch (1), Coffin (5), and Moos and Koslin (14), influence being positively related to the degree of ambiguity.

mean response for Noncommittal *S*s with the mean post-*P* response for the *S*s in both Committal Groups, it is clear that the Noncommittal and GD Committal Groups differ only on the initial exposure and on Block 5. The *S*s in the ID Committal Group, however, tend to give increasingly higher judgments over the course of the five blocks—despite the fact that on the initial exposure the mean response is essentially the same as that of the GD Committal Group.

Superficially, then, it seems that after the initial interaction, both Noncommittal and GD Committal Groups show essentially the same degree of susceptibility to influence when the distance is relatively small and remains constant over a series of interactions. Does this mean that the influence *process* is essentially the same in both groups?

Some light is shed on this question by examining Fig. 1B, in which the mean pre-*P* judgments are presented for the two Committal Groups along with the Control Group means. Any differences between these means would indicate that statistical interaction effects are occurring in the Committal Groups. From Fig. 1B it is quite clear that such interaction effects do exist, as there are wide differences among the three groups on all blocks. It is seen that, on the initial exposure, the mean pre-*P* response of the GD Committal Group (similar to the ID Committal Group) is virtually identical with the Control Group mean—this serves as a sampling check, since there should be no significant difference between these means if the groups were drawn from the same population. Aside from the first exposure, however, consistently higher judgments are being given by the GD Committal Group, on all five blocks. The behavior of the ID Committal Group is similar but even more marked, showing a distinct rise over the five blocks.

Thus, it is clear that a considerable amount of influence is being reflected in the pre-*P* (Committal) judgments of the two Committal Groups. One might think that, since each exposure was perceived as discrete, the effects of *P*'s influence attempts would appear only in the difference between post- and pre-*P* responses on any particular exposure or block. The data indicate, however, that in a continuous committal situation as employed by this study, the *intertrial* influence effects are considerably greater than the *intratrial* effects. Figure 2 shows that the Committal Groups tend to change relatively little from pre- to post-*P* (intratrial), although once the interaction pattern has been established there is seen a large amount of change in the pre-*P* responses (relative to what would be expected in the absence of *any* prior influence attempts). Figure 2A shows that, in the GD Committal Group, the pre-*P* estimates gradually increase in magnitude until, by Blocks 3 and 4, *S* is frequently anticipating *P*'s estimate to an exposure; thus there often is little or no discrepancy (i.e., no "distance") between *S*'s pre-*P* response and *P*'s estimate. In fact, not only do some *S*s "beat *P* to the punch," but a small number of *S*s occasionally give pre-*P* judgments in Block 4 which are higher than *P*'s. Consequently, less pressure is being placed on these committal *S*s than on the Noncommittal *S*s. Therefore, the fact that there is minimal change from pre- to post-*P* judgments in the GD Committal Group does not necessarily indicate a general resistance to influence as a result of self-committal (as suggested by Table 1).

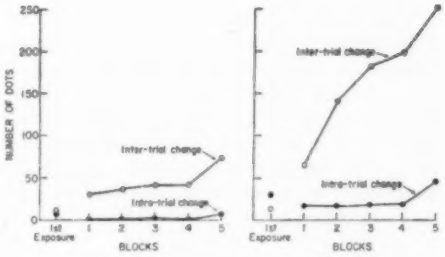

FIG. 2A FIG. 2B

FIG. 2A. COMPARISON OF INTER- AND INTRATRIAL CHANGE IN GD COMMITTAL GROUP
FIG. 2B. COMPARISON OF INTER- AND INTRATRIAL CHANGE IN ID COMMITTAL GROUP

However, this situation does not obtain in the ID Committal Group, since *P*'s estimate is *always* higher than *S*'s. The *S* can never successfully anticipate *P*'s estimate, and there is a constant pressure maintained on all exposures during the first four blocks, and even greater pressure in Block 5. Nevertheless, Fig. 2B indicates that the intratrial change, while somewhat greater than seen in the GD Committal Group, is still quite negligible when compared to the intertrial change. It seems clear, then, that there is a genuine and marked tendency toward resistance observed in Com-

mittal *S*s. On the basis of these results, however, the expression "resistance to change" should be employed judiciously inasmuch as an individual may be highly resistant to intratrial change (i.e., once he has committed himself, he tends to stick to his original judgment) and yet be susceptible to a considerable amount of social influence over the course of an interaction series.

DISCUSSION

Committal vs. Noncommittal

The results of this study suggest a number of interesting aspects occurring over and during the course of a continuous committal situation. On the initial interaction, there is a tendency for less influence to appear in *S*s who commit themselves immediately prior to an attempt at influence. On subsequent interactions, the absolute change from the committal response to the postinfluence response is consistently small. The major effects are seen in a carry-over process which produces marked change in the committal responses themselves.

There is the important question of what factors in the committal process are the effective determiners of the observed resistance to intratrial change. Hobart and Hovland (9) speak of "public commitment," suggesting that the salient property is the revealing of one's response to others. In the present study, it is possible that asking the *S*s to "pay no attention" to *P*'s estimates constituted an important factor. Or, it may be that *S* feels a need to "save face" in the sense that once he had made a response he cannot be talked out of it, although he will allow himself to be influenced considerably in making his actual committal response. This sort of behavior is strikingly similar to behavior noted in a hypnotic situation (8), where an *S* will readily produce an expected response, but once there is overt expression of the response, *S* will go to great lengths to "save face" when challenged. Another possibility is that, in a committal situation, *S* has the opportunity to crystallize his perception of the stimulus. There is no sure way of choosing between these alternatives at the present time, although data from Deutsch and Gerard (6)—while not directly applicable—suggest that "private commitment" yields more influence than "public commitment." The complexity of the problem is emphasized when Kelley and Volkart's results (11) are considered: they were unable to show that resistance to change is greater under public conditions.

Intertrial Social Influence

It is our belief that the "intertrial" phenomenon is an important one, with a number of significant implications. It is probable, however, that the appearance of this phenomenon is limited to certain conditions. It should be noted, for example, that in our experimental design, *P*'s influence responses were consistently *higher* than *S*'s. This permitted *S*s to acquire reliable expectations about *P*'s behavior. Thus, it is apparent that while *S*s may be unwilling to change very much a response already made, they do attempt to anticipate and adjust to *P*'s behavior. This is seen most clearly in the ID Committal Group, where *P*'s response was dependent upon *S*'s committal response: in this situation, *S* could never successfully adjust to *P*, although the ever-increasing committal responses indicate *S*'s attempt to "catch up" to *P*, in between influence attempts.[5]

On a priori grounds, one would expect that no intertrial influence would be reflected if *S* cannot perceive some pattern in *P*'s behavior. For example, if *P*'s responses were randomly distributed in direction (i.e., some higher, some lower, than *S*'s response), *S* would have no opportunity to anticipate *P*. It seems likely, however, that in any situation where *P* communicates a bias which obtains for a number of different issues, *S* will tend to respond to influence attempts both within and between interactions.

A more recent study, in which the intertrial phenomenon was observed under somewhat more subtle conditions, will be reported in the near future. There is also an interesting suggestion of the intertrial phenomenon in a study of Mausner (12), where a confederate attempted to influence *S*'s judgments on the Meier Art Judgment Test. On one half of the judgments (odd), the confederate gave his oral

[5] Whether there is a genuine perceptual change occurring in *S*s or whether their responses merely constitute some form of "public compliance" (cf. 7) cannot be definiteiy answered at the present time. Extrapolating from Asch (2), Berenda (3), and Hoffman (10), however, it is probable that at least some *S*s who are greatly overestimating on the late exposures were relatively unaware of their error, suggesting that a certain degree of perceptual reorganization had occurred.

judgment first; on the other half (even), *S* went first. Mausner's Table 1 (p. 391) shows that as the prestige of the confederate was increased (by manipulation of instructions) *S*s tended to produce more errors *on the even trials* (i.e., when *S*'s judgment preceded the partner's).

Quantitative Measurement of Social Influence

Suppose one is interested in estimating how much influence has occurred following an interaction. One approach consists of determining the degree to which *S*'s (or a group's) post-*P* response differs from what would have been expected in the absence of any attempted influence. This "expected response" may have been ascertained by taking a pre-*P* measurement, or may have been obtained through the use of norms. The quantitative difference between the judgment measured *after* the influence attempt and the expected response gives an estimate of how much "movement" has occurred.

For example, if we knew somehow that *S* would have responded with 100 in the absence of influence, but we find that after *P* said 500, *S* responds with 200, it can be said that *S* "moved" 100 units. Or, in an attitude and opinion study using a 7-point scale of agreement (where −3 is extreme disagreement and +3 complete agreement), if *S* rated an item as −2 (pre) and after influence rated the same item as 0 (post), *S* has moved two units. Movement, however, is defined independently of *P*'s judgment, and consequently is independent of distance; that is, if *P* had said 5000 (rather than 500) and *S* then said 200, *S*'s movement would still be 100. Or, if *S* had first rated an item as +1 and later as +3, *S*'s movement would again be two units.

A measure which reflects the amount of movement *relative to the distance S might have moved* can be called a "Conformity" (C) score. Thus, C is the ratio of movement to distance, and its general formula can be expressed:

$$C = \frac{\text{Post-}P\text{ minus expected response}}{P\text{'s judgment minus expected response}}$$

In the preceding perceptual example, when *P* says 500, C is .25; when *P* says 5000, however, if *S* moved the same 100 units, C is only .02. On the attitude item, if *S*'s prerating was −2 and his postrating was 0, his C score would be 2/5, or .40; however, if his pre- and

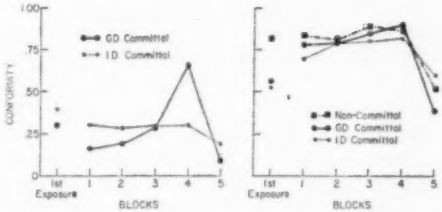

FIG. 3A FIG. 3B

FIG. 3A. CONFORMITY SCORES (C_i) BASED ON INTRATRIAL INFLUENCE
FIG. 3B. CONFORMITY SCORES (C_t) BASED ON ACCUMULATIVE INTERTRIAL INFLUENCE

postratings were +1 and +3 respectively, his C score is 2/2, or 1.00. Note how this measure takes into account a common objection to measuring shifts in opinion (cf. 1), since a shift from −2 to 0 does not mean the same psychologically as a shift from +1 to +3.

Now, in measuring Conformity in any *continuous* interaction series, the question arises as to which "expected response" should be used in the formula: *S*'s pre-*P* judgment or some other estimate, such as norms. Obviously, so long as the two estimates are equal it makes no difference which value is used. We have already seen, however, that in our Committal groups the pre-*P* response, after the initial interaction, is *not* equal to the Control Group norms; hence, it is apparent that different C scores will be obtained depending upon how the "expected response" is determined.

These C scores for our three groups are shown in Fig. 3. Using the pre-*P* response in Fig. 3A, the Conformity scores indicate the extent of *P*'s influence upon *S* *only on the particular interaction or block involved*, ignoring the possible intertrial influence effects which may have preceded; consequently, these C_i (Conformity within a single interaction) scores are markedly attenuated.[6] When Control Group judgments are used (Fig. 3B), however, the *over-all amount of influence* accumulated from the onset of interaction is reflected. It is particularly interesting to note that the C_t (total Conformity) scores are essentially the same in *both* Committal Groups and the

[6] The atypical high C_i score on Block 4 in the GD Committal Group results from the fact that 6 of the 10 exposures in the block elicited mean pre-*P* values equal to or greater than *P*'s estimates (i.e., *S* was successfully anticipating *P*'s response); in such cases, where the distance was zero or negative, C_i was arbitrarily calculated as 1.00, thus spuriously raising the C_i scores.

Noncommittal Group—despite the fact that many *S*s in the ID Committal Group are being exposed to fabulously high judgments by *P* in Blocks 3, 4, and particularly Block 5. The peculiar suggestion that the ID Committal Group shows a higher C_t score (significance indeterminable) on Block 5 than does the Noncommittal Group, cannot be explained adequately until more knowledge is acquired concerning the relationship between "distance" and conformity.

Cautions in Research

For those investigators who study influence in continuous interaction situations, there is an additional implication from our findings. It has long been known that the very act of measuring a response may alter the response. Similarly, measurement of a response may alter *S*'s subsequent behavior. This effect is particularly possible in research designs in which the investigator requires a pretreatment measure. Solomon (16) and Cantor (4) have pointed out the statistical interactions which may occur.

Unless the complete history of the interactions is known, large errors can be made in estimating the amount of influence occurring over a series of interactions. For example, if *S* and *P* have been interacting for 25 "trials" about which we have no data, we may be interested in measuring *P*'s influence on *S* on the 26th trial. We might well obtain a preinfluence measure (*S*'s committal response on the 26th trial) and then a postinfluence measure. When a Movement or Conformity score is now computed, it may be found that *P* had little influence on *S*. However, if during the first 25 trials *P* has been communicating relatively consistent expectations to *S*, *P*'s overall influence may have been considerable, although unmeasurable—since the influence effects may be reflected not in the difference between post- and preinfluence measures on the 26th trial, but in the preinfluence (committal) measure itself.

SUMMARY AND CONCLUSIONS

In a situation in which *S*s are requested to make continuous, discrete perceptual judgments in the presence of a "role-player" attempting to influence *S*, it was found that (*a*) when the distance between the "role-player's" judgment and *S*'s judgment is small, *S*s who commit themselves immediately prior to each influence attempt show more resistance on the initial interaction than do *S*s who do not commit themselves; (*b*) when an abrupt shift is made to a larger distance, more influence is seen again in noncommittal *S*s, and on the remaining interactions at this distance the noncommittal *S*s show a consistent tendency to be influenced to a greater degree than committal *S*s; (*c*) an analysis of two groups of *S*s who committed themselves immediately prior to each influence attempt indicates that the major influence effects in these *S*s appear *between* interactions, in a process which reveals marked change in the committal responses themselves. Thus, immediate self-committal seems to produce "resistance to change" only in the sense that, aside from the initial interaction, there is relatively little change from the committal response to the postinfluence response. Once an *S* has expressed a judgment to a stimulus, that is, he is generally reluctant to change his response to any great degree. The *S* may still be heavily influenced in the situation, however, the effects of which are reflected in subsequent committal responses.

In order to reflect these different processes adequately, it was shown that at least two measures of social influence must be differentiated: one which will give an index of the amount of change within a single interaction, and one which will represent the over-all accumulative effects from the onset of interaction.

REFERENCES

1. ASCH, S. E. The doctrine of suggestion, prestige, and imitation in social psychology. *Psychol. Rev.*, 1948, **55**, 250–276.
2. ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
3. BERENDA, RUTH W. *The influence of the group on the judgments of children*. New York: King's Crown Press, 1950.
4. CANTOR, R. R. The use of extended control-group designs in human relations studies. *Psychol. Bull.*, 1951, **48**, 340–347.
5. COFFIN, T. E. Some conditions of suggestion and suggestibility. *Psychol. Monogr.*, 1941, **46**, No. 4. (Whole No. 241.)
6. DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 629–636.

7. FESTINGER, L. In *Group relations at the crossroads* (M. Sherif & M. O. Wilson, Eds.) New York: Harper, 1953. Pp. 232–256.
8. FISHER, S. The role of expectancy in the performance of posthypnotic behavior. *J. abnorm. soc. Psychol.*, 1954, **49**, 503–507.
9. HOBART, E. M., & HOVLAND, C. I. The effect of "commitment" on opinion change following communication. *Amer. Psychologist*, 1954, **9**, 394. (Abstract)
10. HOFFMAN, M. L. Some psychodynamic factors in compulsive conformity. *J. abnorm. soc. Psychol.*, 1953, **48**, 383–393.
11. KELLEY, H. H., & VOLKART, E. H. The resistance to change of group-anchored attitudes. *Amer. sociol. Rev.*, 1952, **17**, 453–465.
12. MAUSNER, B. Studies in social interaction: III. Effect of variation in one partner's prestige on the interaction of observer pairs. *J. appl. Psychol.*, 1953, **37**, 391–393.
13. MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
14. MOOS, M., & KOSLIN, B. Prestige suggestion and political leadership. *Publ. Opin. Quart.*, 1952, **16**, 77–93.
15. SHERIF, M., & CANTRIL, H. *The psychology of ego-involvements*. New York: Wiley, 1947.
16. SOLOMON, R. L. An extension of the control group design. *Psychol. Bull.*, 1949, **46**, 137–150.

Received March 16, 1955.

ABSTRACT THINKING AND SOCIAL BEHAVIOR IN SCHIZOPHRENIA[1]

JOHN H. FLAVELL

Clark University[2]

THEORISTS have long asserted that schizophrenics differ from normals in their use and understanding of words (1, 6, 9, 10, 11, 12). One aspect of the schizophrenic's disorder is said to be a relative impairment in the ability to focus upon, comprehend, and utilize the abstract and essential rather than concrete and superficial meanings of verbal symbols. Most experimental studies of schizophrenic and normal language behavior tend to support this view (4, 5, 7). Cameron (3) and Sullivan (8) have further asserted that the language disorder is intimately related to the schizophrenic's social withdrawal and general disorganization of interpersonal relationships. Cameron, particularly has stressed the fact that language disorganization breeds social isolation, since effective interaction with others depends in large measure upon an intact linguistic organization.

The purpose of the present study is to investigate experimentally the schizophrenic's supposed difficulty in dealing with word meanings and to attempt to find out whether a relationship between this impairment and impairment of social relationships can be demonstrated. Two general hypotheses are tested: (*a*) In a special test situation in which *S* is repeatedly asked to select the one of two choice words which is "closest in meaning" to a given stimulus word, normals select choice words bearing an abstract, essential meaning relationship to the stimulus words more often than schizophrenics; (*b*) within the schizophrenic group, the ability or tendency to select "abstract" choice words on this test is positively correlated with adequacy of everyday social interaction as rated by judges.

[1] The present study is based on a dissertation submitted to the Graduate School of Clark University in partial fulfillment of the requirements for the degree of Doctor of Philos[illegible]. The writer wishes to express his gratitude for the assistance and encouragement of his thesis committee: Drs. Heinz Werner, Seymour Wapner, John Bell, and Richard Lazarus.

[2] Now at University of Rochester.

METHOD

Subjects

The *S*s were 20 normal and 24 schizophrenic males. The normals were psychiatric aides who responded to a hospital-wide call for volunteer *S*s. The experimental *S*s were chronic schizophrenics whose total hospitalizations ranged from one to twelve years. Although none of the schizophrenic *S*s was in complete clinical remission, there was considerable variation in degree of egointactness. Some were barely testable and some showed no apparent cognitive impairment whatsoever. The two groups were very closely matched for education and verbal intelligence as measured by the Wechsler-Bellevue Vocabulary subtest. The groups were not matched for age, although all *S*s were between 20 and 45. The mean ages of normal and schizophrenic *S*s were 29.2 and 35.2 respectively.

Apparatus and Procedure

Test of Word Meanings. The test material consisted of 56 cards. On each card was printed in capitals a single word at the top (stimulus word) and two words side by side at the bottom (choice words). All stimulus words were nouns; the choice words were variously nouns, adjectives, and verbs. On a given card, the stimulus word was related to one of the choice words in one of eight different ways and related to the other choice word in a different one of these eight ways. The eight types of relationships between stimulus word and choice word were as follows:

1. *Synonyms* (*Syn*). The choice word is a synonym or substitute term for the stimulus word, e.g., "car"—"auto."

2. *Supraordinate* (*Supra*). The choice word refers to a class or genus of events subsuming the object referred to by the stimulus word, e.g., "dog"—"animal."

3. *Subordinate* (*Sub*). The choice word refers to a member of the class which the stimulus word designates, e.g., "month"—"April."

4. *Part-Whole* (*PW*). The choice word refers to a part of the object to which the stimulus word refers, e.g., "deer"—"antlers."

5. *Adjective* (*A*). The choice word refers to a quality or attribute commonly possessed by the object to which the stimulus word refers, e.g., "tea"—"hot."

6. *Verb* (*V*). The choice word refers to an action commonly performed on or by the object referred to by the stimulus word, e.g., "dog"—"bark."

7. *Concrete Context* (*CC*). The choice word refers to a setting or context in which the object referred to by the stimulus word is commonly found, e.g., "streetcar"—"city."

8. *Homonym* (*H*). The choice word is physically similar to the stimulus word but is not related to it in meaning, e.g., "boy"—"buoy."

Of the above eight categories, the first three are considered abstract, essential meaning relationships; the next four are concrete, nonessential meaning relationships; the last is a concrete, physical relationship in which the meanings of the words play no central part. Each of the above eight types of relationship was paired twice with each other type of relationship. Thus, the 56 test cards were composed as follows: stimulus word—*Syn*, *Supra*; stimulus word—*Supra*, *Syn* (order of choice words reversed in the replication); stimulus word—*Syn*, *Sub*; stimulus word—*Sub*, *Syn*, etc., until all possible pairs of relationships had been represented twice. Sixteen additional cards containing choices between *Syn* and words bearing no relation to the stimulus word were included in the test as control words. Since, as will be seen, the *S*s' task involved choices based upon meaning relationships, any individual who showed a definite tendency to choose "no relation" words was assumed not to be task oriented and was eliminated as an *S*.

The *S*s were tested individually. Each was told that he would be shown cards with one word at the top and two at the bottom and that his task would be to choose, from the two bottom words, the word which he felt was the closer in meaning to the word at the top. After three sample cards were shown, the *S* was successively confronted with the series of cards and his choices recorded. The 72 cards (56 test and 16 control) were presented in a random order which was constant for all *S*s. An *S*'s score for a particular relationship consisted simply of the number of times he had selected a choice word exemplifying that relationship. Thus, for example, if an individual chose a *Syn* 14 out of the 16 times *Syn* appeared as a possible choice, his *Syn* score would be 14. Each *S* was given a score for each of the eight relationships and, in addition, was given an over-all *Abstract* (*Abst*) score, which was simply the sum of scores on those relationships designated as abstract and essential, i.e., *Syn*, *Supra*, and *Sub*.

Ratings of social behavior.[3] All members of the schizophrenic group were rated on four traits which, on a priori grounds, were considered to be among those essential to effective social interaction. These traits are summarized as follows:

1. *Sociability:* The extent to which the *S* spontaneously tries to interact with those about him, regardless of the success of his efforts.

2. *Emotionality:* The extent to which the *S* responds with appropriate emotionality to people and events about him; a patient with either flat affect or inappropriate affect would be rated low on this trait.

3. *Coherence:* The extent to which the *S*'s communications are coherent or understandable to those for whom they are intended; the fact that what is clearly communicated might be delusional in content does not detract from an *S*'s score on this trait.

4. *Awareness:* The extent to which the patient gives evidence that he pays attention to or seems aware of the happenings going on about him.

The schizophrenic *S*s were rated on each trait by means of a four-point scale. A rating of "four" on *Sociability*, for example, would mean that the *S* in question exhibits normal or near normal sociability in his day-to-day behavior; a rating of "one" would indicate a virtual absence of spontaneous sociability. A patient's over-all score for social interaction was simply the sum of the separate trait scores. Senior psychiatric nurses, thoroughly familiar with the social behavior of the patients under their charges, served as judges in the study. Since the 24 schizophrenic *S*s were drawn from six hospital buildings, each judge rated only those patients on her building. The *E* explained to each nurse individually and in detail the meaning of the traits to be rated, cautioning her to avoid "halo effect" in her ratings, to rate each *S* in comparison with schizophrenic patients in general rather than just with those on her building, etc. In order to determine interjudge reliability of the scale, three nurses independently rated ten patients selected randomly from those under their care. The mean of the intercorrelations among judges' ratings was .83.

RESULTS AND DISCUSSION

A statistical analysis of group differences on meaning relationship scores is shown in Table 1.[4] The significant group difference in mean *Abst* scores appears to confirm Hypothesis *a* above, i.e., that schizophrenics have relative difficulty in grasping, as essential, those aspects of a word's meaning structure which are the more abstract, unvarying ones. The term *relative difficulty* is used advisedly, however, since the two distributions show considerable overlap. In fact, 33 per cent of the patients achieved *Abst* scores above the normal group mean, and 25 per cent of the normals scored below the schizophrenic group mean. Although useful in formal experimental investigations such as the present one, the present test of word meanings has obvious limits as a clinical instrument for differential diagnosis. It will be noted that, although only the *H* group difference actually attains statistical reliability, the normals consistently exceed the schizophrenics on all three individual abstract relationships and are surpassed by the patients in all but *A* of the nonessential relationships. The fact that *H* choices appear to be the almost exclusive prerogative of certain *S*s in the schizophrenic group is in accord with both Baker's experimental results (2) and the clini-

[3] The author will supply copies of the test of word meanings and the social behavior rating scale upon written request.

[4] The statistical analyses reported in Table 1 were made by means of the Mann-Whitney *U* Test with the exception of the *H* comparison which was analyzed by Fisher's Exact Test in terms of presence or absence of *H* choices in each *S*'s protocol. One of 20 normals made at least one *H* choice as against 8 of 24 patients.

TABLE 1

NORMAL AND SCHIZOPHRENIC GROUP SCORES FOR VARIOUS TYPES OF MEANING RELATIONSHIPS

Relationship	Mean			*SD*	
	Normal	Schiz.	Difference	Normal	Schiz.
Abst	33.05	30.49	+2.56**	2.72	3.70
Syn	12.95	11.83	+1.12	.86	1.97
Supra	11.30	10.58	+.72	1.14	1.78
Sub	8.80	8.08	+.72	2.01	2.27
PW	6.65	7.67	−1.02	1.80	1.95
A	6.40	5.87	+.53	1.56	1.88
V	5.15	5.29	−.14	1.85	1.81
CC	4.70	5.75	−1.05	1.68	1.92
II	.01	.92	−.91*	.44	2.43

* Significant at .05 level or better (two-tailed test).
** Significant at .01 level or better (two-tailed test).

cal observation that some, but not all, schizophrenic patients exhibit "clang association" in their everyday verbal behavior. It is a matter for speculation whether the normal group's order of preference among relationships shown in Table 1, i.e., *Syn*, *Supra*, *Sub*, *PW*, etc., would represent that of the general population of adult thinkers. It is reasonable to suppose that one generates a sort of "habit-family hierarchy" of meaning responses when one apperceives words, a hierarchy which has as one dimension the essentialness-nonessentialness of meaning. Further research is needed, both with respect to the structure of the meaning organizations which are habitually imposed upon words and with respect to the ontogenetic development of these organizations.

The mean and *SD* of the social interaction scores for the patients were 10.83 and 2.27, respectively. The hypothesized positive relationship within the schizophrenic group between adequacy of social interaction and ability to relate words on an abstract basis appears to be confirmed by the data. The rank-order correlation between *Abst* scores and social behavior scale scores is .60, statistically significant beyond the .01 level of confidence. It is clear, however, that such a finding needs further experimental corroboration before Cameron's hypothesis can be said to have been validated. The fact that different patients were rated by different judges is only one of the crudities of the measure of social interaction used in this study. Even if an overall positive relationship between social disorder and language disorder in schizophrenia should be firmly established by further research, there is much more to learn. Neither language disorder nor social disorder is a unitary phenomenon, and it is important to find out how certain types of the one interact with certain types of the other. For example, the communications of a certain voluble but deteriorated schizophrenic appear to have a deleterious effect on his social integration beyond the actual communicative inadequacy involved. Other patients actively seek to avoid the verbal productions of this "crazy talker," thus engendering, vicious-circle fashion, further social isolation and, presumably, further entrenchment into autistic language. Such observations suggest tempting but as yet unexplored research problems.

SUMMARY

The experiment was designed to test two hypotheses: (*a*) normals will exceed schizophrenics in the ability to select, as most similar in meaning to a given word, that word which is related to it in an essential, abstract way; and (*b*) within a schizophrenic group, the above ability will be positively correlated with adequacy of everyday social interaction. Twenty normal and 24 schizophrenic *S*s were presented a series of cards, each bearing a stimulus word and two response words variously related in meaning to the stimulus word. The *S*s were asked to select the response word which they felt to be "closer in meaning" to the stimulus word. Hypothesis *a* was confirmed (.01 level) by group differences in the number of response words selected which bore an abstract relationship to their respective stimulus words. Hypothesis *b* was likewise confirmed by a rank-order correlation of .60 (.01 level) between number of "abstract" response words selected and adequacy of social relationships as measured by ward nurses' ratings.

REFERENCES

1. ARIETI, S. Special logic of schizophrenic and other types of autistic thought. *Psychiatry*, 1948, **11**, 325–338.
2. BAKER, R. W. The acquisition of verbal concepts in schizophrenia: a developmental approach to the study of disturbed language behavior. Unpublished doctor's dissertation, Clark Univer., 1953.
3. CAMERON, N. Experimental analysis of schizophrenic thinking. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley, Calif.: Univer. of Calif. Press, 1944. Pp. 50–64.

4. CHODORKOFF, B., & MUSSEN, P. Qualitative aspects of the vocabulary responses of normals and schizophrenics. *J. consult. Psychol.*, 1952, **16,** 43–48.
5. FEIFEL, H. Qualitative differences in the vocabulary responses of normals and abnormals. *Genet. Psychol. Monogr.*, 1949, **39,** 151–204.
6. GOLDSTEIN, K. Methodological approach to the study of schizophrenic thought disorder. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley, Calif.: Univer. of Calif. Press, 1944. Pp. 17–40.
7. MORAN, L. J. Vocabulary knowledge and usage among normal and schizophrenic subjects. *Psychol. Monogr.*, 1953, **67,** No. 20 (Whole No. 370).
8. SULLIVAN, H. S. The language of schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley, Calif.: Univer. of Calif. Press, 1944. Pp. 4–16.
9. VIGOTSKY, L. S. Thought in schizophrenia. *Arch. Neurol. Psychiat.*, 1934, **31,** 1063–1077.
10. WERNER, H. *Comparative psychology of mental development.* Chicago: Follett, 1948.
11. WHITE, W. A. The language of schizophrenia. *Arch. Neurol. Psychiat.*, 1926, **16,** 395–413.
12. YACORZYNSKI, G. K. An evaluation of the postulates underlying the Babcock Deterioration Test. *Psychol. Rev.*, 1941, **48,** 261–267.

Received March 23, 1955.

SELF-CONSISTENCY AS A FACTOR AFFECTING IMMEDIATE RECALL[1]

DESMOND S. CARTWRIGHT

University of Chicago

THIS experiment was designed to examine two related hypotheses of the client-centered theory of personality. Rogers states:

> As experiences occur in the life of the individual, they are either (a) symbolized, perceived, and organized into some relationship to the self, (b) ignored because there is no perceived relationship to the self-structure, (c) denied symbolization or given a distorted symbolization because the experience is inconsistent with the structure of the self (11, p. 503).

This proposition is based largely upon observation during therapy hours. In translating it for experimental examination, substitution was made of the word *stimuli* for the word *experiences*, and of the words *immediately recalled* for the words *symbolized*, etc. Verbal stimuli were selected as being susceptible of operational evaluation of their (consistent or inconsistent) relations with the self-structure. A consistent stimulus is defined as having a content or meaning for the individual that is related to or descriptive of some aspect of his self-structure. An inconsistent stimulus is defined as one having a content or meaning that is either unrelated to or descriptive of the opposite of some aspect of his self-structure. The resulting null hypothesis to be tested is that there is no difference in immediate recall performance between consistent and inconsistent stimuli.

Closely related to the first proposition are two others made by Rogers:

> Any experience which is inconsistent with the organization or structure of self may be perceived as a threat, and the more of these perceptions there are, the more rigidly the self-structure is organized to maintain itself (11, p. 515).
>
> Psychological maladjustment exists when the organism denies to awareness significant sensory and visceral experiences, which consequently are not symbolized and organized into the gestalt of the self-structure (11, p. 510).

The implication of these two statements would seem to be that if there is a difference in recall performance between consistent and inconsistent stimuli, such difference should be greater for an individual whose self-structure is "more rigidly organized to maintain itself," namely a person who is "psychologically maladjusted." Contingent upon rejection of the first hypothesis then, the second null hypothesis to be tested in this study is that the difference in recall performance between consistent and inconsistent stimuli is the same for adjusted and maladjusted groups.

METHOD

Subjects. The sample of 80 *S*s was heterogeneous as to sex, age, intelligence, and student or nonstudent status. Statistical control was retained on these variables. The total sample was stratified for adjustment however. Persons applying for psychotherapy and persons rated by counselors as therapeutic failures (1–5, on a 9-point scale) are known to be poorly adjusted by comparison with persons rated as therapeutic successes (6–9, on the 9-point scale) and with matched persons who have not had or do not intend to have therapeutic aid (12). The *S*s of the experiment were made up of these four classes as follows: 25 pretherapy, 20 success, 10 failure, and 25 no-therapy *S*s. The latter were taken mainly from the control groups of ongoing researches at the Counseling Center, University of Chicago, and included students, their spouses, and staff members of the University. The a priori stratification was checked by two independent criteria of adjustment, the Q Adjustment Index (2), and Counselor Ratings (12).

Procedure. The first part of the experiment was aimed at establishing for individual *S*s which stimuli were consistent for them and which inconsistent.

Fifty-six cards, 2 x 4 inches, were prepared, bearing self-descriptive statements like "I have an attractive personality." The statements were taken from the Q sort developed by Butler and Haigh (1) The cards were to be sorted by *S* into seven piles of eight, each pile being placed under an indicator card. There were seven indicator cards, reading, for example, "Pile 1; most unlike me," or "Pile 7; most like me."

The introductory statement made by *E* was as follows:

> We are doing some studies on memory, and later on I shall be asking you to do a number of different things involving memory. Some of these things may seem like tests you have taken before, of intelligence, and so on. But they are being used

[1] This paper is based upon the author's dissertation, "A Study of Imbalance in Immediate Recall," submitted in partial fulfillment of requirements for the doctoral degree at the University of Chicago, Department of Psychology, 1954. The author is indebted to his sponsoring committee, John M. Butler, Carl R. Rogers, and William F. Soskin. Grateful acknowledgment is made to the Research Fellowship Committee, of the Counseling Center, University of Chicago, for assistance by research fellowship funds and for permission to use clients and controls as subjects in this experiment.

here for purely scientific purposes of investigating memory function. We ask you to do your best on these tests, but want you to know that we are making no evaluation of yourself. Before we start, I want you to sort some cards. Here are 56 cards with statements on them. Sort the cards to describe yourself as you see yourself today, from least like you in Pile 1, to most like you in Pile 7.

The *S*s were allowed as long as they needed to sort the 56 cards. When they were through, *E* checked that the pile 7 cards were most like *S*, and then placed each pile in a pigeon-hole of a tray concealed from *S*. The *E* then presented *S* with a sheet bearing a column of 20 nonsense syllables taken from Glaze's list of syllables with 53 per cent association value (7). The *S* was also given a dittoed sheet bearing two parallel columns of 20 lines and spaces. The left column was headed "List"; the right column was headed "Your own." Pencils were provided. The *E* then gave the following instructions:

Here are 20 nonsense syllables. As you can see, each one is made up of three letters, a consonant, a vowel, and another consonant. Each one can be pronounced, but it has no meaning. Look at them carefully, and write each one down here (under "List"). Then make up another 20 such syllables of your own, printing them down here (under "Your own"). Be sure to make each one with a consonant, a vowel, and then another consonant.

While *S* wrote syllables, *E* examined the Q sort cards and scored for adjustment. Each statement had been selected for its unambiguity and its facility of description by a single adjective. Fifty-six cards, 1 x 2 inches, were prepared, each bearing an adjective comparable with a self statement. For the most part the adjective was contained exactly in the statement. For "I have an attractive personality," the adjective was "attractive." The *E* selected the adjectives appropriate to the cards in *S*'s piles 1, 2, 6, and 7 (least like and most like). For each pile the adjectives were divided into two lots of four. Four from pile 6 were shuffled face down with four from pile 7. This set of eight constituted the first list of consistent adjectives to be used in subsequent testing. Four from each of piles 1 and 2 constituted the first list of inconsistent adjectives. The second lists of each kind were made up similarly.

When *S* had completed writing the syllables, the sheets were removed by *E*. Then *E* gave *S* a set of 35 cards, 1 x 2 inches, each bearing the name of a common object. These names were selected from a list of 140 names of objects that persons might feel interested or involved in owning. The 35 names were all of undramatic items like sneakers, rabbits, sailboat, guitar, chessmen, paintings, television, typewriter, etc. Seven indicator cards were set out before *S*, each bearing a pile number, and ranging from "Pile 1; least like me to own" to "Pile 7; most like me to own." The *S* was instructed as follows:

I want you to sort these cards on the basis of whether you feel it would be like you to own them. If it would be most like you to own a particular one, place it in pile 7, and if it would be least like you to own something, place that in pile 1, and arrange the rest accordingly in between. Sort the cards into seven piles, with five cards in each pile. You have about five minutes for this. OK?

While *S* sorted the objects, *E* prepared the score sheet for the syllables. Two lists of eight syllables each were to be used in the test. Each list contained four Glaze and four *S* syllables, (those *S* had made up himself) alternately arranged, the first list beginning with a Glaze syllable, the second beginning with an *S* syllable. The Glaze syllables were already printed in position on the score sheets. The *E* filled in the appropriate spaces with *S* syllables. The syllables were selected on the basis of serial position within the column of syllables *S* had written as his own. The serial positions of Glaze and *S* syllables were matched.

When *S* finished his sort of the objects, *E* checked that pile 7 contained the "Most like me to own" ones. The *E* then collected the piles in the order 7 and 1, 6 and 2, 5 and 3, and lastly 4. The collection procedure was aimed at eliminating differential recency effects for the test to follow. Having placed each pile in its own pigeon-hole number, *E* presented *S* with a blank sheet of paper and a pencil, and said:

Now I want you to write down as many as you can remember of the objects you just sorted; just as many as you can remember in two minutes.

After two minutes, checked on a stop watch accurate to ⅕ sec., the sheet was withdrawn. Later it was scored by checking with *S*'s card sort and entering the pile number against each word that was a 1, 2, 6, or 7. The sum of *S*'s 6 and 7 entries was his score for consistent stimuli on this test. The sum of 1 and 2 entries was the score for inconsistent stimuli. In later sections these scores will be referred to as scores on the Possessions Test.

Following this test, *E* took the score sheet for the syllables and instructed *S* as follows:

I am going to read out a list of nonsense syllables. I want you to remember as many as you can, and tell them back to me when I have finished. All right?

The first list of 8 syllables was then read out by *E* as distinctly as possible, using one pitch of voice and at the rate of one per second. All vowels were taken as short. Labials were emphasized. The consonants C, K, and Q were spoken equivalently as K. The maximum recall time was 30 seconds. As *S* said back the syllables, *E* checked them on the score sheet. After the second list had been read and checked, *E* counted the number of *S* syllables correctly recalled in both lists. This number was *S*'s score for consistent stimuli on this test. The number of Glaze syllables correctly recalled in both lists constituted *S*'s score for inconsistent stimuli.

Briefly, the rationale for the consistency values in regard to nonsense syllables was that syllables made up by *S* himself would carry some small amount of personal meaning, that is, be slightly consistent, whereas the Glaze syllables would not.[2]

In future reference this test will be called the Nonsylls Test.

After the Nonsylls test, *E* took the first pile of con-

[2] Clearly the Glaze syllables are neutral, the kind of stimuli *S* would ignore in ordinary life. They are called "inconsistent stimuli" here only for the purpose of maintaining linguistic uniformity.

sistent adjective cards, face down, and also the score sheet. Preliminary work on this test had shown five trials to be necessary, so that the score sheet was arranged in a 4 × 5 table. Instructions to *S* were as follows:

> I am going to read some adjectives which correspond to the statements you used to describe yourself. I want you to try and remember them in the order I read them out, and when I finish you say them back to me in that order. Don't worry if you can't say them all; just say back as many as you can in the same order. OK? For repeat trials *E* said "Let's try that list again." For remaining lists *E* said "Here's another list."

Following the instructions the first consistent list was read out by *E* in one pitch at the rate of one per second. As each card was turned up and read out, it was placed on the score sheet, concealed from *S*, until all 8 cards were plainly visible to *E* and in their appropriate order for repeat trials. As *S* recalled the list, *E* counted the number correct in order starting from the first. This number was entered in the appropriate list row and trial column of the score sheet. Trials were repeated until all 8 adjectives were recalled in correct order, or until the end of the fifth trial. Where a list was correctly recalled before the fifth trial, the remaining trials were scored 8, and the next list begun straight away.

To balance practice, fatigue, and proactive inhibition, the first consistent list was followed by two inconsistent lists, the second consistent list coming last. Scores were summed across trials for both consistent lists to give the score for consistent stimuli on this test. Sums across trials for both inconsistent lists gave the score for inconsistent stimuli on this test. In future, reference will be made to this test as the Adjectives test.

The three experimental tests, Nonsylls, Possessions, and Adjectives, may be seen to differ both in type of materials and in instrumental methods. The former were varied to cover a range of classes of stimuli. The latter were varied in order to reduce the possibility of finding spuriously comparable results on all tests due to similarity of instruments.

Testing was continued after the Adjectives test for the examination of an hypothesis unrelated to those at issue in the present report. This subsequent testing is important here only in that it included measures of intelligence and digit memory.

RESULTS

Table 1 shows means and standard deviations for the two adjustment groups on consistent and inconsistent stimuli for the three experimental tests. The success and no-therapy *S*s are the adjusted group. The pretherapy and failure *S*s are the maladjusted group. With heterogenous subjects and groups, the appropriate test for the first hypothesis is the Sign Test. This was applied to the differences in recall performance between consistent and inconsistent stimuli for the total sample of 80 *S*s. Results for Nonsylls were 23/46 (zeros omitted); for Possessions 13/56; for Adjectives 23/55. All results were in favor of consistent stimuli, and all have a confidence level of better than .005 (one-tailed). Nineteen *S*s favored consistent stimuli on all three tests; 42 *S*s favored consistent stimuli on two tests; 9 *S*s favored inconsistent stimuli on two tests; 1 *S* favored inconsistent stimuli on all three tests. Adjustment groups contributed approximately equally to these results.

TABLE 1

SUMMARY OF CONSISTENT AND INCONSISTENT RECALL SCORES

Test	Adjusted Group				Maladjusted Group			
	Consistent		Inconsistent		Consistent		Inconsistent	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Nonsylls	3.47	1.10	2.93	1.07	3.63	1.12	2.71	1.43
Possessions	6.78	1.46	5.44	1.76	6.66	1.22	4.49	2.24
Adjectives	43.86	13.89	38.98	13.75	37.77	14.33	31.45	15.00

Note.—Scores are based on 45 adjusted and 35 maladjusted *S*s.

Before reaching any conclusions about the first hypothesis of this study, it is necessary to determine whether or not the foregoing results may be accounted for by differential response-availabilities associated with word frequencies (8). In relation to the Nonsylls test, a similar hypothesis is possible; namely that syllables made up by *S*s may more closely resemble words than do those presented by *E*.[3]

The syllables recalled by the 46 *S*s who favored their own syllables were sought among Glaze's lists (7). More than 83 per cent were found. Mean association-values for own syllables recalled were then estimated for each of the 46 *S*s. These values were compared with those for the Glaze syllables (all 53 per cent). The result was 16/30 ($p < .10$) in favor of own syllables. Results for adjustment groups were: adjusted 8/16 ($p < .25$); maladjusted 8/14 ($p < .25$). The indication is that part of the superiority in recall for own (consistent) syllables may be attributed to differences in association-value, presumably equivalent to closeness of resemblance to words. While many *S*s recalled own syllables having 0 per cent or other values considerably less than 53 per cent,

[3] This hypothesis was suggested in a critical reading of this paper. The subsequent examination shows that future research on this test should have syllables of higher association-value than 53 per cent presented by *E*.

it is clear that at best the present results may be contaminated with a tied variable.

In regard to the Possessions test, the 35 names of objects were assigned relative frequency values from the Thorndike-Lorge count (13). On this basis they were split into seven groups of five each and further assigned numbers from 1–7 accordingly. The preliminary hypothesis tested was that there are no differences in recall between the 10 most frequent and 10 least frequent words. Results for 80 *S*s were 19/45 ($p < .01$); for the adjusted group 9/30 ($p < .01$); for the maladjusted group 10/15 ($p > .25$). The null hypothesis is rejected for the total sample and for the adjusted group, but not, curiously, for the maladjusted group. For the total sample then, recall for more frequent stimuli was superior to recall for less frequent stimuli. For frequency differences to account for the results on consistency, however, two things must hold: (*a*) *S* recalls the more frequent words either better than or as well as the less frequent words; and (*b*) the mean frequency value of the consistent stimuli is greater than that of the inconsistent stimuli. The mean frequency values of consistent and inconsistent stimuli were computed for each of the 56 *S*s who favored consistent stimuli in recall. Assigning a plus where the consistency difference may be accounted for by frequency differences, and a minus where the reverse holds, the Sign Test was applied. Results for 56 *S*s were 27/29; for the adjusted group 13/17; for the maladjusted group 12/14. All results are clearly nonsignificant. It is evident that the results on the Possessions test cannot be accounted for in terms of a word-frequency hypothesis.

The hypothesis tested in regard to the Adjectives test was that there is no difference in recall between the two lists of greater mean frequency and the two of lesser mean frequency. Frequency values were taken from the Thorndike-Lorge count. The adjective lists of a 50 per cent sample, 20 adjusted and 20 maladjusted *S*s, were used. Results were 15/24 ($p < .25$) in favor of the lists having lesser mean frequency values. Results for the adjusted group were 9/10, not significant; for the maladjusted group 5/15 ($p < .05$). As would be expected from these results, the further hypothesis that consistent lists do not differ from inconsistent lists in terms of mean frequency value cannot be rejected. The result for this comparison was 19/21. Clearly differences in word-frequency cannot account for the results on the Adjectives test in terms of consistency differences.

While results on the Nonsylls test are equivocal, then, the experimental findings on the Possessions and Adjectives tests do warrant rejection of the first null hypothesis of this study.

The second hypothesis to be tested in this study was that the difference in recall performance between consistent stimuli and inconsistent stimuli is the same for adjusted and maladjusted groups.

The relation of adjustment to the consistency variable was examined in two ways. First, ratio scores for all subjects on each test were computed by subtracting the score on the inconsistent stimuli from the score on consistent stimuli and dividing by the sum of the two scores. The mean ratio was computed for each group, no-therapy, success, pretherapy, and failure. For each of the three tests these mean ratio scores were ranked in order of size. A *W* coefficient (9) was then computed on these rankings and found to be .911, significant at better than the .01 level for a one-tailed test. In all cases the mean ratio scores of the maladjusted *S*s were greater than those for the adjusted *S*s.

This result gives gross support for the hypothesis that a greater difference between recall for consistent and for inconsistent material is associated with a lesser degree of adjustment.

It may be argued that the experimental results may be due to differences in rote-memory ability or concentration, which is known to be associated with adjustment status (14). Similar questions may be raised in regard to sex, age, intelligence, or status differences among groups. Ratio scores were not found to be affected by these variables however.

For the purpose of obtaining refined evidence on the relation between adjustment and the effects of the consistency variable, a sample of 40 *S*s was drawn from the experimental population in the following manner. On the Wechsler test for digit memory (14) the failure *S*s ranged in score from 7 to 16. Ten *S*s from each of the other three groups were selected to match on digit memory scores, giving a total of 40 *S*s in the sample, with 20 adjusted and 20 maladjusted. (The matching

was exact except for two pairs: adjusted 9 and 10, maladjusted 7 and 8, respectively.) These two groups did not differ in sex, age, intelligence and status composition.

Though the primary stratification of groups on adjustment is inferred from their relationship to or outcomes of therapy, independent criteria were obtained. The Q adjustment scores of all *S*s in the selected sample were transformed to logarithms. A two-way analysis of variance was computed, testing for adjustment, previous experience of therapy, and the interaction of these, against the within-groups mean square. The only significant value of F was that for adjustment, $F = 3.932$, which for a one-tailed test, and 1 and 36 degrees of freedom, is significant at better than the .05 level of confidence. However, this measure is known to be subject to error due to defensive sorting (3). For the success, pretherapy and failure *S*s counselor ratings of adjustment on a nine-point scale were obtained. The t test between the 10 success *S*s (adjusted) and 18 (two not rated) maladjusted *S*s was significant at better than the .001 level for a one-tailed test ($t = 4.15$, with 26 degrees of freedom).

So far as the relevant factors have been accounted for, it seems reasonable to assume that the adjusted group of 20 and the maladjusted group of 20, matched for digit memory, do not differ in any respect save adjustment.

The following examination will deal only with the Possessions and the Adjectives tests, since results on the Nonsylls test were seen to be open to some doubt. Scores for the 40 *S*s on the Possessions test were transformed to logarithms to take account of the smallness of score entries (4). In Table 2, the separate analyses of variance for the consistent and inconsistent stimuli are summarized.

The interpretation of the results for the Possessions test would seem to be that the difference between recall performance for consistent stimuli and for inconsistent stimuli is a function of adjustment. It appears too that the difference between adjustment groups is due to reduced efficiency in the recall performance of the maladjusted group on the inconsistent stimuli.

Analyses of variance for the complete Adjectives test yielded no significant F ratios. It appeared from inspection of the data, however, that this result might be due to serial effects over the four lists of adjectives. Analyses of variance for the first lists separately yielded the results shown in Table 3.

The interpretations of the results for the Adjectives test (first lists) would seem to be closely similar to those for the Possessions test. It will be recalled that five trials were allowed for each list of adjectives. For each adjustment group the mean number of ad-

TABLE 2

ANALYSIS OF VARIANCE OF RECALL SCORES ON THE POSSESSIONS TEST

Source of Variation	*df*	Mean Square	*F*	*p*
Consistent possessions				
Adjustment	1	.00521	—	
Therapy	1	.00944	1.42	>.20
Interaction	1	.01691	2.55	>.10
Digit levels	9	.00491	—	
Residual within groups	27	.00664		
Total	39			
Inconsistent possessions				
Adjustment	1	.57081	9.08	<.005*
Therapy	1	.00716	—	
Interaction	1	.05867	—	
Digit levels	9	.03750	—	
Residual within groups	27	.06290		
Total	39			

Note.—Scores are transformed to logarithms.

* p value for a one-tailed test.

TABLE 3

ANALYSIS OF VARIANCE OF RECALL SCORES ON THE ADJECTIVES TEST (FIRST LISTS)

Source of Variation	*df*	Mean Square	*F*	*p*
Consistent adjectives				
Adjustment	1	22.400	—	
Therapy	1	144.200	1.675	>.20
Interaction	1	5.200	—	
Digit levels	9	41.600	—	
Residual within groups	27	86.081		
Total	39			
Inconsistent adjectives				
Adjustment	1	330.575	6.206	<.025*
Therapy	1	34.175	—	
Interaction	1	55.325	1.039	>.20
Digit levels	9	60.792	1.141	>.20
Residual within groups	27	53.266		
Total	39			

* p value for a one-tailed test.

jectives recalled in correct order per trial was computed for all lists separately. While a drastic reduction in recall performance occurred on the first list of inconsistent adjectives for the maladjusted *S*s, on the second list of inconsistent adjectives, they seem to have recovered somewhat, while the adjusted group continued to show a decrement in performance level. This is a curious result and no explanation can be offered at the present time.

The results of the ratio-score analysis, of the Possessions test, and of the first lists of the Adjectives test would seem to warrant rejection of the second null hypothesis of this study. Results on the second lists of the Adjectives test, however, require that some reservation be made.

DISCUSSION

The group data of this experiment would seem to support Rogers' position that experiences which are inconsistent with the structure of self may be denied symbolization or given a distorted symbolization. There was considerable evidence of distortion in immediate recall for inconsistent adjectives. Typical examples, all from different *S*s are: "satisfied" was misrecalled as "dissatisfied" for one trial; "hopeless" was misrecalled as "hopeful" for two trials; "hostile" was misrecalled as "hospitable" for three trials; "tolerant" was misrecalled as "intolerant" for five trials. Since *S*s had sorted their cards to say they were *not* satisfied, hopeless, etc., it is clear that misrecall resulted in making the stimulus word consistent with their self-structure. It seems likely that much fruitful research can be done in this area.

Since maladjusted *S*s were found to show weaker recall than adjusted *S*s for inconsistent stimuli, it seems reasonable to say that inconsistent experiences may be threatening, as hypothesized by Rogers. Interpreted in this way, the present results would seem to confirm, under conditions involving no experimental threat variable, the findings of other workers (5, 6, 10) in regard to "defensive forgetting" under induced ego-involvement.

SUMMARY

This study examined two related hypotheses derived from Rogers' theory of personality. The first null hypothesis tested was that there is no difference in immediate recall performance between stimuli that are consistent with the self-structure and stimuli that are inconsistent. The second hypothesis was that the difference in recall performance, if any, between consistent and inconsistent stimuli is the same for adjusted and maladjusted groups.

The 80 *S*s in this study were heterogeneous in age, sex, intelligence, and status. Three different tasks of immediate recall were used, involving nonsense syllables, names of common objects, and self-descriptive adjectives. Consistent syllables were those *S* had previously made up himself. Consistent names of objects were those *S* had stated to be most like him to own. Consistent adjectives were those appropriate to self-descriptive statements *S* had asserted to be most like himself.

On all three tasks consistent stimuli were recalled significantly better than inconsistent stimuli. Findings on the nonsense syllables, however, were open to the doubt that some part of the results may be due to differential resemblance to words. A differential word-frequency interpretation of results on the other two tests could not be sustained. The first null hypothesis of this study was therefore rejected.

Two approaches were made to the second null hypothesis. An over-all test involving all *S*s on all three tasks suggested that the difference in recall between consistent and inconsistent stimuli is a function of adjustment. Refined evidence was obtained from the data on objects and adjectives for 40 *S*s stratified for adjustment and matched for rote memory. In regard to the objects and the first lists of adjectives the difference between adjustment groups was found to be due to reduced efficiency in the recall performance of the maladjusted group on the inconsistent stimuli. Though with some reservation due to findings on the second lists of adjectives which showed certain unexplainable serial effects, the evidence seems to be in favor of rejecting the second null hypothesis.

Discussion centered upon evidence of distortion in immediate recall and upon the relation between the present results and those of other workers on "defensive forgetting."

REFERENCES

1. BUTLER, J. M., & HAIGH, G. V. Changes in the relation between self-concepts and ideal concepts consequent upon client-centered counseling. In C. R. Rogers, & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954. Pp. 55–75.
2. DYMOND, ROSALIND F. An adjustment score for *Q* sorts. *J. consult. Psychol.*, 1953, **17**, 339–342.
3. DYMOND, ROSALIND F. Adjustment changes in the absence of psychotherapy. *J. consult. Psychol.*, 1955, **19**, 103–107.
4. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
5. ERIKSEN, C. W. Defense against ego-threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, **47**, 230–235.
6. ERIKSEN, C. W. Individual differences in defensive forgetting. *J. exp. Psychol.*, 1952, **44**, 442–446.
7. GLAZE, J. A. The association value of non-sense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.
8. HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229–234.
9. KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.
10. LAZARUS, R. S., & LONGO, N. The consistency of psychological defense against threat. *J. abnorm. soc. Psychol.*, 1953, **48**, 495–499.
11. ROGERS, C. R. *Client-centered therapy*., Boston: Houghton Mifflin, 1951.
12. ROGERS, C. R., & DYMOND, ROSALIND F. (Eds.) *Psychotherapy and personality change*., Chicago: Univer. of Chicago Press, 1954.
13. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Bureau of Publications, Columbia Univer., 1944.
14. WECHSLER, D. *The measurement of adult intelligence* (3rd Ed.). Baltimore: Williams and Wilkins, 1944.

Received March 28, 1955.

SOCIAL STATUS AND SUCCESS STRIVINGS[1]

ELIZABETH DOUVAN

Survey Research Center

SYSTEMATIC differences between child-training procedures employed in the two major social classes have been reported by several investigators (3, 6, 7). These differences cluster in two related areas of training: development of internalized controls and learning of achievement motivation. Davis (3) has noted the critical role of anxiety as a reinforcement for achievement strivings in the middle class training scheme, and Ericson's study (7) reveals a picture of the environment of the middle class child as making early and consistent demands for personal attainment.

Since middle class society not only places great stress on accomplishment, but imposes demands earlier than does the working class (7), we would expect the need for achievement to be more generalized in middle class children than in children of lower status.[2] Success-failure cues in any situation should, among middle class children, elicit a relatively consistent reaction, irrespective of the reward conditions of the specific situation. Since working class children are taught achievement strivings neither so early nor so systematically their reactions to success-failure cues should be more responsive to changes in the reward potential of the situation in which such cues occur.

The present study was designed to contrast members of the two social groups with respect to the degree of achievement motivation they manifest in two success-failure situations which differ in reward potential. In one situation success was defined as achieving an abstract norm; in the other successful performance offered, in addition, a material reward. The specific hypothesis to be tested was that working class youth would manifest a significantly greater difference in achievement strivings under the two reward conditions than would youngsters from the middle class.

METHOD

Subjects

The *S*s for this experiment were adolescents in the senior year of high school in a medium-sized midwestern community. Of the original group of 336, 23 were excluded from the analysis because of incompletions in at least one of the measures used in the study.

School children were used as *S*s because the schools provide established groups of individuals from both social classes under similar objective conditions. In addition *S*s of this age were especially suitable because of the likelihood of finding among them a substantial proportion who had been subject to the influences of their class subcultures during the crucial early training period.

The schools in which the study was done provided a good cultural setting for an investigation of class differences. The social situation in these schools is characterized by relatively good apparent chances for mobility and by strong pressures exerted toward mobility by teachers and by the middle class satisfactions reflected in the well-kept and expensive homes surrounding the schools. Working class students who do not yield to such pressures—who identify with the working class even in this situation—should be stable representatives of people with working class values and goals.

Measures and Controls

Class membership. Two types of indices were used in defining *S*s' class position. The first was an occupational index based on data derived from school records and a questionnaire in which *S* was asked to describe his father's work and conditions of employment. The second index was a subjective class assignment based on a variation of Centers' technique.[3] Only those *S*s for whom the two indices agreed on the middle or working class were used in the test of the study's central hypothesis.

Independent placements of each *S* on the basis of father's occupation were made by two judges, with a percentage agreement of .95. The criteria for assigning occupations to the two classes are listed below.[4]

[1] This article is based on a doctoral dissertation submitted to the University of Michigan. The author is indebted to Theodore M. Newcomb, Max L. Hutt, and other members of her doctoral committee for their able guidance.

[2] The importance of time of training has been highlighted by McClelland (10). He hypothesizes that behavior patterns learned in the preverbal period will be particularly generalized and intractable since the original learning cues are unavailable to consciousness.

[3] In addition to the Centers' question, and preceding it, *S*s were asked: "Which one of the following social classes (middle, lower, working, upper) would you like to be a member of?" and "Which one . . . do you think you'll be a member of in ten years?" By including these questions, it was hoped to reduce elements of fantasy and aspiration in *S*s' answers to the question on present class membership.

[4] The rationale for this classification system is presented fully in the original report from which this article is drawn (5). Placements made on the basis of this scheme have been shown to correlate highly with those based on Warner's more complex system (12).

Middle Class

a. Addresses self to people.
b. Exerts control over people either in a disciplinary capacity within the business power unit, or as an active agent of control over people outside the business unit. In either case the exertion of control serves the interest of the business unit.
c. Income can be increased through individual efforts—some degree of effect on the market for his services and skills.

Working Class

a. Addresses self to tools.
b. Exerts control over no one in the job situation.
c. Income relatively fixed. No control over market for his labor, the only source of his income.
d. Easily replaceable skill.

Most occupations were relatively easily placed by means of these criteria. If a job met two of the three criteria for the middle class, it was assigned this position. Most commonly discrepancies occurred between the first criterion and later ones. Thus a surgeon might be said to address himself primarily to tools, but he exerts control over others (patients, nurses, secretarial help) and has some power over the market for his skills. The factory set-up man represents the opposite case of this discrepancy. Fewer than five per cent of the occupations rated revealed any such inharmonious elements.

Need Achievement. McClelland, Atkinson, and coworkers have developed a method for measuring achievement motivation based on a special scoring of an adapted TAT. Their measure, which made possible an experimental test of the hypothesis of the present study, is presented in detail in a recent publication (11).

Scoring was done by two judges, trained by a psychologist experienced in the use of the system. A number of indices of agreement were obtained between each judge and the instructor and between the two judges. Percentage agreements on scoring specific categories ranged from .83 to .96; eight rank-order correlations between total scores assigned by two judges to twenty-five *S*s all yielded values above .91. Score-rescore reliabilities computed on 150 protocols were .87 for one judge and .94 for the other.

Procedure

The *S*s were assigned by a system of random numbers to one of the two motivation conditions. Limitation of funds for material rewards dictated the decision to distribute the sample two-to-one in favor of the symbolic reward condition.

The *S*s were given tests, tasks, and questionnairer in the following order:

a. questionnaire on subjective class identification
b. anagrams "test"
c. motor "test"
d. Need Achievement measure
e. questionnaire on father's occupation

The experiment was introduced in the following manner in all groups:

> A group of psychologists is administering a series of tests to high school students in different parts of the country. These tests tell us about various abilities of people. We're going to give the tests to you now. I'll describe the ability each test measures . . . before we give it. After each test I'll give you the average score of high school students on that test, so you'll know how you're doing as we go along.

In the material reward condition this introduction was followed by a short statement concerning the fact that we had found that students often did not do as well as they might because they didn't always try hard. For this reason, they were told, ten dollars would be given to any student whose over-all score on the tests reached a certain value:

> We'll announce the winning scores tomorrow, but in the meantime, you'll have the average scores of students we've already tested in (*name of city*) so you'll know how you're doing.

The effectiveness of this device for increasing motivation was attested by the excitement which followed the announcement. A questionnaire asking each *S* to estimate the relative magnitude of his own and others' desire for the prize corroborated the impression of the *E*s.

The anagrams and motor tasks were used to create failure experiences and thus establish a nonsatiated state of the achievement motive. Before each of the "tests," *S*s were told what it measured and how much time they would be given. They were admonished against any discussion during the testing period, and also against "giving up" if they did poorly on early tests. Since the important thing was the over-all score, success on later tests might compensate for humble beginnings.

Following a 4-minute trial on each "test," a falsely high average was announced for "high school students in your city," and we guided *S*s rapidly into the next experimental step.

The *n* Achievement "test" was dissociated from the two previous ones. A different *E* administered it, and *S*s were told that this test would not be counted in their scores. It was introduced as a test of creative imagination, and the instructions conformed closely to those used by McClelland *et al.* (11).

At the end of the testing period it was necessary to relieve anxieties created by the experimentally induced failure. The previously announced high school norms were inaccessibly high, and our desire was to substitute for them more reasonable standards which would leave *S*s with the feeling that they had performed well relative to their peers. To effect this substitution, a minor drama was staged.

An experimental assistant came into the room near the close of the hour and hurriedly whispered something to *E*s. Following this communication, one *E* announced with obvious embarrassment that because of a last-minute rush in preparations, he had by mistake been given supplies intended for the team of psychologists giving the tests at the local college. The norms he had previously announced for the tests had not, therefore, been the scores of a high school group, but were scores obtained by college students. He now had the correct high school norms, and would read them to *S*s.

These belated but "correct" norms were quite moderate, having been established on the basis of pretest experience to be within the near-reach of even the slow high school senior. A free question and discussion period followed and concluded the experiment.

RESULTS

Summary data for tests of hypotheses are presented in Table 1. There is, we see, a significantly greater difference between the *n* Achievement scores of working class *S*s under the two reward conditions than between those of middle class *S*s. This finding confirms our hypothesis that middle class youngsters would manifest more generalized achievement strivings, thus showing themselves more impervious to immediate situational demands.

One might ask whether the greater relative shift in striving among working class children does not simply reflect the fact that money is scarcer in their environment and therefore more highly desirable to them. The material reward offered might, according to this argument, be sufficient to trigger additional motivation in the working class child but insufficient to bring into play in the middle class youth any comparable increment of striving.

Results of the questionnaire on the desirability of the ten dollar reward oppose this interpretation—there was no class difference in conscious desire for the reward. The sum of ten dollars was used specifically to guard against differential subjective responses to the material reward.

The *n* Achievement scores of the class groups under the two reward conditions can be viewed in another way. The fact that under the material reward condition there is substantially no difference in the mean scores of the two class groups ($t = .27$; $p > .30$), indicates that the measure of *n* Achievement as such is not seriously class biased. We might consider the condition of material reward a control situation designed to establish the applicability of the *n* Achievement measure to both groups. On this interpretation, results in the situation characterized by the absence of material reward can be generalized thus: In response to a failure, defined as not achieving an abstract standard, middle class individuals manifest a significantly higher need to achieve than working class individuals ($t = 4.3$; $p < .01$).

Since each *S* knew only his own scores, success did not entail any immediate gratification of prestige strivings or needs for recognition from the group. This extreme limitation of the satisfactions to be gained by success in the nonreward situation leads to the conclusion that the manifestation of a high need to achieve under such conditions attests to the independent and central nature of success as a motivating factor in the individuals responding in this manner.

TABLE 1

ACHIEVEMENT MOTIVATION OF WORKING CLASS, MIDDLE CLASS, AND MARGINAL *S*s UNDER TWO REWARD CONDITIONS

Class	Material Reward						D_1-D_2	SE of Diff.	t	p
	Present			Absent						
	N	*M*	*SD*	*N*	M	*SD*				
Working	34	8.06	3.61	53	4.89	2.89	2.43*	1.05	2.31	.02
Middle	37	8.30	3.71	73	7.56	3.76				
Marginal	36	7.85	3.72	71	7.71	4.93	.60†	1.14	.53	.25

* Here D_1 refers to the mean difference for working class *S*s under the reward conditions, D_2 to the mean difference for middle class *S*s.

† In this case D_1 refers to the mean difference for middle class *S*s, D_2 to mean difference for marginals.

Marginality

The results reported to this point were obtained when social class was defined on the basis of both objective occupational criteria and a subjective measure. Possibly the most provocative groups defined by this dual scheme are those comprising individuals whose self-assigned social class differs from their objective position in the social structure. From the present analysis it is not possible to make even tentative remarks about downward identifying middle class *S*s, since there were too few of these cases. Consider, however, upward identifying working class *S*s.

Because of the variety of factors which may underlie such a misidentification (e.g., ambiguity in the status structure, simple ignorance, status anxiety, defensive resistance against one's actual position), no specific predictions were made about the relative achievement motivation of this group under the two reward conditions. However, several general possibilities were explored.

Steiner (13) found a tendency for these marginal cases to behave like middle class individuals, but more extremely. The present study posed two questions relating to this observation: (*a*) Do marginals obtain higher *n* Achievement scores than middle class *S*s under low reward conditions? and (*b*) Do marginals react to the two reward conditions

with more consistent achievement motivation than middle class *Ss*?

This analysis did not yield any significant findings. While the mean *n* Achievement for the marginal group under the symbolic reward condition is higher than the mean of the middle class group (Table 1), the difference is not reliable ($t = .20$; $p > .40$). Nor do the marginals show greater consistency in response to the two situations. There was substantially no difference between middle class and marginal *Ss* with respect to the relative shift in achievement strivings ($t = .53$; $p > .25$).

Another inquiry into the behavior of this group followed from Davis' distinction (3) between original socialization and adaptive acculturation, and his description of the differential demands and effects of these two kinds of learning. Since the marginal *Ss* were probably not uniformly socialized in the middle class value pattern, wide variation was anticipated in the extent and accuracy of their incorporation of middle class behavior patterns. Some might have overlearned, some not yet have learned the accepted middle class responses to success-failure cues in a symbolic reward situation. Greater heterogeneity should therefore be manifested in the responses of this group under the condition of absence of material reward.

The findings confirm this expectation. The standard deviation of the marginal group was not compared to that of the working class group, since their means were too different to allow the assumption that the groups derived from the same universe. But the test which could be made—between middle class and marginal variances—reveals that the marginal group is significantly less homogeneous than the middle class group (F ratio $= 1.6$; $p < .05$).

DISCUSSION

The extent and nature of achievement responses have been shown to depend on social class membership. More autonomous and generalized success strivings characterize members of the middle class, while achievement motivation of working class individuals is more highly dependent on the reward loading of the task situation.

The present results, in conjunction with observations made by others (2, 4, 8, 9, 14), suggest that the nature of achievement motivation among adolescents in the two classes is functional to the dominant values and behavior expectations of the class subcultures. The occupational role of the middle class adult requires a high degree of competitive performance, the product is individual, and responsibility for success and failure is personal. To meet these demands successfully the individual must be equipped with generalized and stable internal motivation to achieve. The industrial manual worker, on the other hand, is more familiar with nonpersonal causality and the effect of external factors (e.g., layoffs) on individual goal attainment. His labor contributes to a group product, and the value of personal competitiveness is minimal in his occupational role. Success and failure are less highly personalized so there is little need to strive unless success involves some meaningful and apparent reward (4).

A caveat should be entered against overinterpretation of these findings. They are not concerned with the way in which achievement striving fits into the total psychological structure of the individual. Thus while Davis (3) attributes to anxiety a major role in the determination of success strivings in the middle class, the present data do not permit conclusions about the kind of learning process which underlies them. Nor are assertions justified about the relative rationality, efficiency, or rigidity of the two patterns of behavior observed, since the psychological economics of individual *Ss* have not been studied. In any case Brown (1) has pointed to the danger that terms like efficiency may serve as screens for evaluation.

The conclusion is warranted, however, that achievement motivation is conditioned by and appropriate to the cultural context in which it develops. The middle class child is urged to individual achievement, is compared to age mates by his parents, and is taught to respond to symbolic as well as material rewards. He develops, accordingly, strong and well-internalized desires for accomplishment, and responds consistently to success-failure cues even when achievement offers little or no substantial reward. The working class child, on the other hand, is not pressed for individual attainment as early or as consistently, and his motivation to succeed in a given task is more clearly related to the rewards such success entails. These differential patterns of response conform to values and life conditions in the

two subcultures and to the behavior expectations which will be imposed when the subjects of this study reach adult status in their respective class cultures.

SUMMARY

Previous research has indicated that the two major social classes differ with respect to norms and values in the area of personal achievement. Middle class parents, in rearing their children, assert demands for individual success earlier and more regularly than do parents in the working class. On this basis it was anticipated that achievement striving would be a more central motivational factor in middle class children, and that their responses to situations containing success-failure cues would be relatively consistent despite variations in rewards offered for success. Since children in working class homes are not so vigorously urged to personal attainment, it was hypothesized that their motivation to succeed would vary more directly with changes in the reward potential of task situations.

High school students from both social classes were given a series of tasks under two reward conditions. In one reward was limited to personal satisfaction derived from attaining a norm; in the other a material reward was added to this satisfaction. Following failure experiences (to induce deprivation) *Ss* in each condition were given McClelland's projective test for achievement motivation.

As predicted, the mean *n* Achievement scores for working class *Ss* under the two reward conditions showed greater variation than did the means for middle class *Ss*. Though members of both class groups responded similarly to the material reward condition, the achievement strivings of working class *Ss* dropped significantly when the material reward was absent, while the motivation of middle class *Ss* remained at approximately the same high level.

The *Ss* who were placed by an objective criterion in the working class, but identified themselves as middle class, were considered marginal and were excluded from the main analysis. It was predicted that the *n* Achievement scores of these *Ss* under the symbolic reward condition would be marked by greater variability than those of other groups. The prediction was confirmed.

From the results it is concluded that the pattern of achievement motivation a child develops depends on the class subculture in which he is trained, and is functional to the values and behavior requirements with which he will be confronted as he assumes adulthood within that setting.

REFERENCES

1. BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469–476.
2. CENTERS, R. Attitude and belief in relation to occupational stratification. *J. soc. Psychol.*, 1948, **27**, 159–185.
3. DAVIS, A. Socialization and adolescent personality. *Adolescence, Forty-third Yearbook, Part I*. Chicago: National Society for the Study of Education, 1944.
4. DAVIS, A. The motivation of the underprivileged worker. In W. F. Whyte (Ed.), *Industry and society*. Chicago: Univer. of Chicago Press, 1946. Pp. 84–106.
5. DOUVAN, ELIZABETH. The influence of social class membership on reactions to failure. Unpublished doctor's dissertation, Univer. of Michigan, 1951.
6. DUVALL, EVELYN M. Conceptions of parenthood. *Amer. J. Sociol.*, 1946, **52**, 193–203.
7. ERICSON, MARTHA C. Social status and child rearing practices. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1947, Pp. 494–501.
8. HOLLINGSHEAD, A. B. *Elmtown's youth*. New York: Wiley, 1949.
9. LYND, R. S., & LYND, HELEN M. *Middletown*. New York: Harcourt, Brace, 1929.
10. MCCLELLAND, D. C. *Personality*. New York: Sloane, 1951.
11. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
12. ORT, R. S. A study of role-conflicts as related to class level. *J. abnorm. soc. Psychol.*, 1952, **47**, 425–432.
13. STEINER, I. A theory and an empirical study of the role of reference groups. Unpublished master's thesis, Univer. of Michigan, 1948.
14. WHYTE, W. F. *Street corner society*. Chicago: Univer. of Chicago Press, 1943.

Received April 6, 1955.

AN EXPERIMENTAL AND THEORETICAL ANALYSIS OF PERCEPTUAL DEFENSE[1]

CHARLES W. ERIKSEN[2] AND C. THAYER BROWNE
Harvard University

WHILE the question of perceptual defense has become the subject of considerable controversy, a firm body of experimental support for such a phenomenon has remained untouched by criticism (5, 10, 13, 14, 22). Most of the controversy has concerned the question of stimulus familiarity and has centered around a few methodologically weak and superficially conceived experiments which were probably unable to test the perceptual defense hypothesis one way or the other. On the other hand, a number of experiments have demonstrated quite convincingly that psychological defense mechanisms as conceived in personality theory are detectable in perceptual recognition behavior. In these experiments not only was care taken to insure that the perceptual stimuli were indeed anxiety arousing for the individual subjects (*S*s), but provision was also made in the experimental design for individual differences in response to anxiety. In addition these particular studies are not open to dismissal on the popular grounds that they failed to control for familiarity or frequency of prior occurrence of the stimuli.

Although these studies have seemingly demonstrated the occurrence of perceptual defense, they have done little to advance our understanding of the processes involved or to relate their findings with other psychological data. Explanations in terms of stimulus familiarity have proven inadequate (18, 27) and concepts such as "response suppression" (29) do little more than substitute one label for another. Probably the most important question concerning perceptual defense, and the one that has provoked the most criticism (19, 24), is how the individual knows what to defend against. It has been argued that the individual to defend against a stimulus must first perceive it, and, therefore, the perceptual defense hypothesis must assume "unconscious perceiving" or "subception" (3, 10, 19, 23, 24). It is the purpose of the present study to show that the perceptual defense hypothesis does not require an assumption of "unconscious perception" and to show further that perceptual defense can be described in terms of behavior theory principles derived from avoidance conditioning and punishment.

The present authors believe that the controversy over perceptual defense and unconscious perception has primarily arisen due to a failure to distinguish between the individual's perception and his response. While operationally it is difficult to separate the two, there are experimental situations where this has been done (8, 20). Failing to distinguish perception from response is probably most confusing in situations where the number or range of available responses is less than the number of perceptual discriminations the individual is capable of making. For example, if an *S* is presented with 11 different-sized squares under conditions of absolute judgment and is allowed to use only two responses such as "large" and "small," we have objective evidence for only two perceptual states existing in the *S*. If, however, we allow him to use 11 different responses on the task, we find that he can discriminate six different size categories with perfect accuracy (7). We can further show that the increase in the number of response categories did not change his perception, for if the S is also instructed to move a lever a distance proportional to the perceived size of the square simultaneously with his verbal response of "large" or "small," we find that the lever-positioning response indicates six reliable perceptual states in our *S* at the same time his verbal response is indicating only two.

A condition of response limitation exists in most, if not all, perceptual defense studies. In need-in-perception experiments using tachistoscopic-like techniques, the *S* is generally instructed, implicitly or explicitly, to guess and to respond in terms of *meaningful* material such as *words* rather than to describe the stimulus "just as he saw it." If the *S* is in-

[1] This research was made possible by a grant from the Laboratory of Social Relations, Harvard University.

[2] Now at The Johns Hopkins University.

structed to respond in terms of whole words, a brief tachistoscopic exposure of the word "shot" may lead to the verbal response "shut." If on the other hand, the *S* is allowed greater freedom in his responses and is instructed to describe just what he perceived, we get data like "I saw that the first letter was an 's' and the second letter was a tall one and it looked like there was a 't' on the end" (2). A somewhat longer stimulus exposure under these instructions might elicit the response "There is an 'sh' followed by a short letter with a 't' on the end." Here we have two different responses indicating two different perceptions, yet if the *S* was instructed to respond only with whole words, both exposures might well have elicited the same response of "shut." If we grant that the responses obtained under instructions requiring the *S* to describe exactly what he perceived more nearly describe the perceived stimulus than do responses obtained under the set to respond with words of whole units, then perceptual defense can be described in a straightforward extension of behavior theory concepts without the bugaboo of "unconscious perception."

The elements for such a behavior theory model are as follows. Let us assume that a response (R_1) has been closely followed by anxiety so that now the cues produced by the occurrence of R_1 are conditioned stimuli for anxiety arousal. Let us further assume that R_1 has been the strongest response (most frequently occurring) in a habit family hierarchy activated by a stimulus (S_1). Since the occurrence of R_1 now leads to anxiety, there will be a change in the relative response strengths of the responses to S_1 (15, 16, 17). Thus S_1 may now tend to elicit R_2 more frequently, a response that was initially lower in the hierarchy than R_1.[3] The extent of this change in response to S_1 would depend, in part, upon the relative reinforcement value of R_1 and R_2. However, R_2 would be expected to gain in relative strength since its occurrence is not followed by anxiety.[4]

This model can be applied to the perceptual recognition behavior of tachistoscopically-presented stimuli. Assume that the word "shot" is being presented to an *S* at progressively increasing exposure durations. At a short duration the *S* might perceive the stimulus as "sh—." Such a stimulus might be expected to evoke a number of different responses such as *sham*, *shut*, *shot*, *shay*, *shed*, *shim*, *shad*, etc. Which of these responses actually occurs would depend upon their relative response strengths which in turn would be dependent upon such variables as prior reinforcements, set, recency and frequency of prior occurrence and punishment. If one of these responses had been previously followed by punishment or anxiety, its probability of occurrence would be expected to be decreased (15, 16, 17). Thus, for example, if the word "shot" led to "self-devaluing" memories and thoughts in our *S*, which in turn evoked anxiety, the probability of *shot* as a response would be expected to be decreased relative to other responses. We would expect, therefore, that our *S* would be less apt to respond correctly at this exposure duration than another *S* for whom the word "shot" did not lead to anxiety provoking memories or associations.

As the duration of the exposure interval increases, the stimulus perceived by the *S* will be expected to change. At a longer interval the *S* might perceive the stimulus "sh–t." This new stimulus would tend to evoke a more restricted number of responses, but here again we would expect the same variables to determine which response was actually evoked. Eventually the exposure interval will become long enough so that the *S* will perceive the entire stimulus "shot" which will have but one response and it will satisfy our criterion for recognition. As can be seen, this model would lead to the prediction that

[3] This is probably the simplest schema of the model. However, the model can be readily extended to cover instances where R_1 is the stimulus for R_2 and R_2 in turn provides the stimulus for still another response, R_3. In this case we have chained conditioning. Now if R_3 should become the conditioned stimulus for anxiety due to being followed by punishment, not only would R_3 be inhibited, but in keeping with stimulus generalization we could expect the inhibition to generalize to R_2 and even to R_1. We might further expect the amount of generalized inhibition to be a function of the degree of punishment that followed R_3.

[4] Although empirically we know that punishment leads to a decrease in frequency of occurrence of the punished response, there is theoretical disagreement as to how anxiety produces the decrease in response occurrence. Our purpose is not to contribute to these theoretical issues but to show that the experimental data on perceptual defense can be handled by the punishment paradigm.

words (or other stimuli) that evoke anxiety would, in a statistical sense, have higher recognition thresholds than neutral words. And, conversely, words associated with positive reinforcement would tend to have lower recognition thresholds.

In the experiment described below we have attempted to test this model of perceptual defense by experimentally manipulating the response strengths for 10 word responses. In one group of *S*s, response strength was increased for these words by manipulating the frequency and recency of prior occurrence of these words. A second group of *S*s also received comparable frequency and recency of experience with the 10 words, but here the experimental arrangement was such as to permit us to assume that these words were also conditioned stimuli for anxiety. It was predicted that the variables of frequency and recency of prior exposure would lower the recognition thresholds for the experienced words, but this gain in recognition would be less in the second group due to the anxiety conditioned to these same words.

Design. Part of the design and some of the experimental techniques used here are similar to those used by Postman and Solomon (26). In the present study two groups of *S*s were administered an anagrams test in an intellectually competitive situation so designed as objectively to produce moderate to severe failure. Upon completion of the anagrams test both groups were told the correct solution to each of the anagrams and required to write these solutions on the back of the test booklet. Following this, the perceptual recognition thresholds for the anagram solutions were obtained along with the recognition thresholds for the same number of comparable control words.

Since consistent individual differences in response to self-esteem threat have been reported (1, 4, 5, 12, 13, 22), it was considered necessary to control for these differences in the present design. This was accomplished by composing one group of *S*s (high group) who scored high on the psychasthenia scale from the MMPI and composing another group of *S*s (low group) who scored low on this scale. Carlson (4) and Eriksen (11) found that *S*s scoring high on the psychasthenia scale showed less impairment in recall of anxiety-related material than low-scoring *S*s, and Eriksen and Davids (9) have further reported that *S*s with high scores on psychasthenia are clinically rated as using repression to a lesser degree than *S*s with low scores. If we examine the individual items comprising the psychasthenia scale and the intercorrelations of this scale with other scales, it would appear that one of the ways that high and low-scoring *S*s differ is in terms of their willingness to admit self-devaluing characteristics (6, 18). The high-scoring *S*s seem especially ready to admit bad things about themselves while the low-scoring *S*s are reluctant to admit any faults. From this observation we have assumed that self-devaluing thoughts and memories are less anxiety arousing for the high-scoring *S*s.

Since recency of occurrence has been shown to increase response strength in terms of the probability of the response being evoked (26), it was expected that the recency of experience with the anagram words would tend to lower their recognition thresholds relative to the control words in both groups. The use of these words in solving the anagrams and writing out the solutions at the *E*'s direction should tend systematically to increase the strength of these words as responses over the control words which had not previously been encountered in the experimental context. Moreover, it was further expected that the high group *S*s would show a greater difference in recognition thresholds between the anagram and the control words than would the low group. Since the anagram solutions had been experienced in a context of failure it was expected that their occurrence as responses would lead to self-devaluing thoughts which, in keeping with our assumption, would evoke more anxiety in the low psychasthenia group. Thus the gain in response strength of the anagram solutions due to recency would be at least partially cancelled in this group due to the stimulus properties of these words as anxiety evokers.

METHOD

There were two experimental sessions. During the first a personality questionnaire containing the *Hysteria*, *Psychasthenia*, *Lie*, and *K* scale items from the MMPI was administered to groups of from one to three *S*s. A month to a month and a half later the second experimental session occurred. It consisted of two phases. In Phase I *S*s were administered the anagrams test and in Phase II their recognition thresholds for the ten anagram solutions and 13 control words were determined. Following this the *S*s were interviewed by the *E* and

questioned about their feelings and performance on the anagrams test. They were then told the nature of the experiment.

The *Ss* were informed that the second session was an experiment on verbal problem solving. The anagrams test was presented in booklet form with one anagram appearing on each sheet and with space on the sheet for the *Ss* to work out their solutions. The anagrams were administered one at a time with a maximum of five minutes allowed for a solution. Within this time limit the *E* manipulated the time allowed the *S* on each anagram so as to determine that no *S* succeeded in completing more than five anagrams. After the time limit was called on each anagram, *S* was instructed to circle his solution. Then the *E* called out the correct answer and the *S* was required to write this down on his answer sheet.

While ostensibly the anagrams were administered to *Ss* in groups of three, actually there was only one real experimental *S* in the group. The other two were stooges of the *E*. Their job was to pretend to solve all the anagrams quickly. This deception was used to increase the experimental *S*'s feeling of failure since the arrangement of the group was such that he could not fail to observe that the other two "*Ss*" were solving *all* the anagrams and *doing it much faster than he*.

Upon completion of the anagrams test, *E* then read aloud again the ten solutions and the *Ss* were required to copy these on the back of their test booklets. This provision was to insure that all ten words were of approximately equal recency in the *S*'s experience. At this point the *E* dismissed the two stooges on the pretext that he had previously scheduled them for a different time for the second phase of the experiment. Then turning to the experimental *S*, *E* said, "You were the subject that I had scheduled to do the second part of this experiment now, weren't you?" The *S* was then asked to wait outside while *E* prepared for the next part of the experiment. A ten-minute wait always occurred before the second phase was begun.

In the second phase the *S*'s visual recognition thresholds were determined for the 10 anagram solutions and for 13 control words. Three of the control words served as practice words to familiarize the *S* with the task and to establish a base level of perceptual performance from which threshold determinations could begin. They also served to minimize variations from practice in the thresholds for the remaining 20 words. The anagrams, the practice, and the control words were the same as those used by Postman and Solomon (26). Each anagram was chosen so as to permit only one solution in terms of a reasonably common English word. All words were seven letters in length and were of relatively low frequency of occurrence as determined by the Thorndike-Lorge Semantic Count (28).

The words were typed in capital letters and mounted as 35 mm. slides. They were projected on a beaded screen from a distance of 20 ft. by a 300-Watt projector with a shutter attachment. All words were exposed at a constant shutter speed of $\frac{1}{100}$ of a second. The visibility of the projected words was varied by changing the voltage across the projector lamp through a Variac in the circuit. A voltage range of 28 to 55 volts was required in the experiment. The ambient illumination on the screen was 2.7 millilamberts from overhead lighting.

Thresholds were determined by the ascending method of limits. After the thresholds for the three practice words had been determined, the thresholds for the remaining words were determined by setting the voltage 6 volts below the lowest threshold value among the three practice words. Starting from this voltage level, three exposures of the first test word were given. If recognition did not occur by the last of these three exposures, the intensity of the projected words was increased by two volts and three more exposures given. Between each exposure, *S* was asked to guess what the stimulus word was. This procedure was followed until correct recognition occurred, each time the voltage being increased by two volts with three exposures given at each voltage level. When correct recognition of a word occurred, the voltage was returned to the initial level and procedure repeated with the next word.

The mean initial voltage for the low group was 35 and for the high group 37. A t test indicated a mean difference of this size could be expected to occur more often than 20 times in 100 by chance alone.

Thresholds were measured as the number of exposures required prior to correct recognition with voltage level being disregarded. If an *S* had failed to recognize a word by the twenty-first exposure (seven voltage levels), the word was shown at maximum intensity and the *S*'s threshold for this word was scored as 21 exposures. This occurred in only a few instances. The order in which the anagram and the control words were presented was random with the restriction that an equal number of anagram and control words occur in each half of the list.

Subjects. Seventy-nine male undergraduate students were hired through the Harvard Student Employment Service to take the personality inventory. The 20 *Ss* with the highest and the 20 *Ss* with the lowest scores on the psychasthenia scale were later contacted and asked to serve in the remainder of the study. Of these 40 *Ss*, 17 were obtained for the high psychasthenia group and 14 for the low group.

RESULTS

For each *S* in both groups, the mean number of exposures required for recognition of the 10 anagram solutions and the 10 control words were separately computed. After Bartlett's test provided no basis for rejecting the assumption of homogeneous variances, these mean values were analyzed via a modified three-way analysis of variance (anagram vs. control words, high vs. low groups, and individuals). The results of this analysis are shown in Table 1.

As can be seen, there is no over-all difference between the high and low groups on our measure of perceptual recognition. There is, however, a significant difference between the anagram and neutral words, and, in addition, there is a significant interaction of the high and low groups with the anagram and the control words. The nature of these sig-

TABLE 1

SUMMARY OF THE ANALYSIS OF VARIANCE

Source	Sum of Squares	*df*	s^2	*F*
Total	890.5	61		
Between high and low groups	48.5	1	48.5	2.11
Between anagram and control words	119.0	1	119.0	14.33**
Individuals	665.4	29	22.9	
High and low groups × anagram and control words interaction	8.3	1	8.3	4.88*
Pooled *Ss* × anagram and control words	49.3	29	1.7	

* Significant beyond the .05 level.
** Significant beyond the .01 level.

TABLE 2

MEAN TRIALS FOR RECOGNITION OF THE ANAGRAM AND CONTROL WORDS BY THE HIGH AND LOW GROUPS

Groups	Words		Combined mean
	Anagram	Control	
High	8.26	11.69	9.98
Low	10.77	12.73	11.75
Combined mean	9.39	12.16	

nificant effects can be seen more clearly in Table 2 where the mean recognition thresholds for the high and low groups on the anagram and control words are shown.

It is apparent from this table that the significant *F* obtained between the anagram and control words is due to the anagram words being more quickly recognized. This finding is consistent with the results obtained by Postman and Solomon (26) and confirms our prediction that the recency of experience with the anagram words would increase the strength of these words as responses.

The significant interaction obtained between the high and low groups with the anagram and control words provides the test of our second hypothesis. In keeping with our assumptions and model, we expected that the increase in response strength due to recency would, at least in part, be cancelled out by anxiety effects associated with these responses in our low group. Thus the high group would be expected to have a lower mean recognition for the anagram words, relative to the control words, than would the low group *Ss*. Examination of Table 2 shows that the significant interaction obtained in the analysis of variance consists of just such a relative difference in recognition thresholds for the anagram and control words between the two groups.

The nature of this interaction can be seen more clearly if for each *S* we subtract his mean recognition threshold for the anagram words from his mean threshold for the control words. If the mean of these difference scores is computed for each group, we find that the high group has a mean difference score of 3.43 exposures and the low group a mean difference of 1.96. A *t* test applied to the difference between these two means yields a value of 2.21 which is significant beyond the .02 level for a one-tailed test.[5]

DISCUSSION

The results of this experiment not only support our hypotheses, but they lend themselves quite well to a description in terms of behavior theory. In keeping with our expectations, the recency of exposure of the anagram solutions strengthened these words as responses to the fragmentary tachistoscopic stimuli. More important, however, was the finding that anxiety associated with these word responses in our low psychasthenia *Ss* counteracted in a large part the increment in response strength these word responses should have gained through recency of experience.

[5] Due to some superficial similarities of the present experimental design to one used by Postman and Solomon (26), it is necessary to point out the reason for discrepancies in the kinds of data analyses used by these investigators and the kinds used here. In the Postman and Solomon study, a comparison of the recognition thresholds for completed and incompleted words on the anagrams test was an essential part of the experiment. In the present study such a comparison would be rather meaningless. The anagrams test was so arranged through the use of stooges as to constitute a failure experience for the test as a whole. While each *S* succeeded in completing some of the anagrams, this achievement could hardly be construed as a success experience since the stooges had obviously completed these anagrams much more quickly than he. Comparison of the recognition thresholds for the completed and incompleted anagrams would also be unfeasible here since a few *Ss* in each group were unable to complete more than three anagrams within the five-minute time limit.

The results of the present experiment could, of course, be explained in other ways and with other concepts. The advantage of the model we have chosen lies in the fact that it enables us to relate the phenomenon of perceptual defense to data and principles derived from research in other areas of human and animal behavior. Dollard and Miller (6) have shown the fruitfulness of an analysis of thought and overt verbal behavior in behavior theory terms and have proposed a theory of repression which is based upon inhibition of thoughts and other responses that are conditioned stimuli for anxiety. The present paper is essentially an extension of their type of analysis to the research on perceptual defense.

One question that our model does not supply an answer for is whether or not the effects of anxiety extend to both overt and implicit verbal responses. All of us know that our overt verbal behavior does not always reflect our implicit verbal behavior and the question arises as to whether the *S* implicitly speaks the response but withholds it from overt speech or does the implicit verbal response also fail to occur? It is not possible to resolve this question satisfactorily at present due to the lack of operational specification of awareness. From the standpoint of behavior theory such a distinction is important only if it can be shown that subsequent behavior differs depending upon whether or not the *S* was subjectively aware of the response. At present little systematic work has been done on specifying awareness and determining the behavioral differences that follow from subjective awareness and unawareness. In some experiments in the perceptual defense area such as that of McGinnies (25) there is reason to believe that the response effects extended only to the overt level. Thus in McGinnies' experiment we might expect the overt speaking of "dirty" words in front of a female experimenter to constitute a stimulus for anxiety, but there is little reason to expect that implicitly speaking the words would produce anxiety. In the present experiment, on the other hand, whatever anxiety was evoked by the response words would occur from implicit speaking of the words. In keeping with the assumptions of our experiment the anagram solutions evoked anxiety because they had become stimuli for self-devaluing thoughts. Thinking the response would be expected to produce as much anxiety as overtly speaking the response. Furthermore once the *S* had thought of the response word, there would appear to be no reason for inhibiting the voicing of the word, particularly since there would appear to be a certain reward value in performing well on the experimental task through achieving rapid recognition. Thus the present experimental design would seem to require the elimination of the thought itself if anxiety was to be successfully avoided.

SUMMARY

The present paper has presented an account of perceptual defense in terms of behavior theory principles derived from research on punishment and avoidance conditioning. It was shown that an assumption of "unconscious perception" was not required to explain the experimental data on perceptual defense if certain distinctions are made between the *S*'s verbal response and his perception.

An experiment was presented that attempted to test the behavior theory model. In one group of *S*s the response strengths for ten word responses were increased by manipulating the frequency and recency of prior occurrence of these words. A second group of *S*s received comparable recency and frequency experience with these ten words, but here the experimental arrangement was such as to permit the assumption that these words would also become conditioned stimuli for anxiety. It was predicted that the variables of frequency and recency of prior exposure would lower the perceptual recognition thresholds for the experienced words, but this reduction in thresholds would be less in the second group due to the anxiety conditioned to these words. The results of the experiment confirmed the predictions.

REFERENCES

1. ATKINSON, J. W. The achievement motive and recall of interrupted and completed tasks. *J. exp. Psychol.*, 1943, **46**, 381–390.
2. BRICKER, P. & CHAPANIS, A. Do incorrectly perceived tachistoscopic stimuli convey some information? *Psychol. Rev.*, 1953, **60**, 181–188.
3. BRUNER, J. S., & POSTMAN, L. An approach to social perception. In W. Dennis, *et al.* (Eds.) *Current trends in social psychology.* Pittsburgh: Univer. of Pittsburgh Press, 1948.
4. CARLSON, V. R. Individual differences in the

recall of word-association-test words. *J. Pers.*, 1954, **23**, 77–87.

5. CHODORKOFF, B. Self-perception, perceptual defense, and adjustment. *J. abnorm. soc. Psychol.*, 1954, **49**, 508–512.
6. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
7. ERIKSEN, C. W. Comparison of the accuracy and interdependence of concurrent verbal and lever positioning responses to size stimuli under absolute judgment conditions. USAF, *WADC Tech. Rep.*, 1955, No. 55–170, in press.
8. ERIKSEN, C. W., & WECHSLER, H. Some effects of experimentally induced anxiety upon discrimination behavior. *J. abnorm. soc. Psychol.*, in press.
9. ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor Anxiety Scale and the hysteria-psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135–137.
10. ERIKSEN, C. W. The case for perceptual defense. *Psychol. Rev.*, 1954, **61**, 175–182.
11. ERIKSEN, C. W. Psychological defenses and "ego-strength" in the recall of completed and incompleted tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 45–50.
12. ERIKSEN, C. W. Individual differences in defensive forgetting. *J. exp. Psychol.*, 1952, **44**, 442–446.
13. ERIKSEN, C. W. Defense against ego-threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, **47**, 230–235.
14. ERIKSEN, C. W. Perceptual defense as a function of unacceptable needs. *J. abnorm. soc. Psychol.*, 1951, **46**, 557–564.
15. ESTES, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, **57**, No. 3 (Whole No. 263).
16. GWINN, G. T. The effects of punishment on acts motivated by fear. *J. exp. Psychol.*, 1949, **39**, 260–269.
17. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.
18. HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229–235.
19. HOWIE, D. Perceptual defense. *Psychol. Rev.*, 1952, **59**, 308–316.
20. HYMAN, R., & HAKE, H. W. Form recognition as a function of the number of forms which can be presented for recognition. USAF, *WADC Tech. Rep.*, 1954, No. 54–164.
21. KERRICK, JEAN S. Some correlates of the Taylor Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1955, **50**, 75–77.
22. LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, **19**, 471–483.
23. LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: A study of subception, *Psychol. Rev.*, 1951, **58**, 113–123.
24. LUCHINS, A. S. On an approach to social perception. *J. Pers.*, 1950, **19**, 64–85.
25. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
26. POSTMAN, L., & SOLOMON, R. L. Perceptual sensitivity to completed and incompleted tasks. *J. Pers.*, 1950, **18**, 347–358.
27. SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43**, 195–202.
28. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers' College, Columbia Univer., 1944.
29. WHITTAKER, EDNA M., GILCHRIST, J. C., & FISCHER, JEAN W. Perceptual defense or response suppression? *J. abnorm. soc. Psychol.*, 1952, **47**, 732–733.

Received April 7, 1955.

OBJECT CONSTANCY IN SCHIZOPHRENIA: THE ENHANCEMENT OF SYMBOLIC OBJECTS AND CONCEPTUAL STABILITY[1]

HAROLD L. RAUSH

University of Michigan

CONSTANCY studies have for the most part dealt with phenomena of size, shape, brightness, and color. As Brunswik (5) has pointed out, however, more complex judgments may also be subsumed under constancy theory. Just as proximal stimuli on the retina tend to reduce the accuracy of size estimations, so any of a number of varying cues may reduce the accuracy of object perception generally. Studies on the influence of needs, values, and stress in perception indicate some conditions under which object constancy becomes modified. Thus, there is evidence that stimuli which have value (1, 4) or which are relatively meaningful (13) tend to be overestimated in size as compared to neutral or relatively meaningless stimuli. The first of the two studies presented here examines such modifications of object perception as a function of the type of stimuli and the ego-structure of the subject, a problem that is implied in the negative findings of other studies (8, 9).

In similar vein, concepts and various cognitive phenomena tend toward constancy. As Krech and Crutchfield state, "Just as it is fortunate that object constancy is the rule in visual perception, so is it fortunate that our more complex cognitive field tends to maintain its *major* organizations despite apparently contradictory facts and experiences. If this were not true, there could be little stability in our cognitive life with respect to our most central beliefs and attitudes." They add, "Stability, of course, can become rigidity; and when that happens . . . we frequently get into trouble with the world of reality . . ." (10, p. 100). The second study investigates the problem of cognitive organization in relation to the individual's ego-structure.

The studies utilized three groups: 30 male paranoid schizophrenic patients, 30 male nonparanoid schizophrenic patients, and 30 male college student controls. The groups are described elsewhere (12); it is sufficient here to say that both patient groups were fairly young, fairly well oriented, and were relatively equivalent to each other in degree of "sickness."

The experimental materials consisted of 35 5-by-8-inch cards. At the center of each card a circle was drawn in black india ink. The circles were arranged in the package of cards in ascending order by size. The smallest circle was 4.6 mm. in diameter, and each succeeding circle was one-twelfth larger in diameter than the circle immediately preceding it. The value of one-twelfth was chosen on a rough empirical basis as representing approximately one jnd. The cards were numbered from 1 through 35 in ascending order. Thus, card 2 is approximately one jnd larger than card 1; similarly, card 35 is approximately one jnd larger than card 34.

STUDY 1: THE ENHANCEMENT OF SYMBOLIC OBJECTS

Procedure. Each subject, who had just taken part in a traditional size constancy experiment (12), was brought into a room with ordinary daylight conditions. He was given the pack of cards and was asked to select for each of the following objects a circle that would represent its size. These objects were (*a*) an average overcoat button, (*b*) an average scoop of ice cream, and (*c*) the thickness of an average cigar. The overcoat button was selected as a meaningful object, rather neutral in affective tone. The other objects were chosen as having more potential affective value from the viewpoint of psychoanalytic theory. Thus, while it is true that any and all objects may be libidinized by a particular person, it was assumed that this potential would be greater for the ice cream and cigar than for the overcoat button. Furthermore, there was no reason to believe that patients and controls would differ with respect to their familiarity with these objects, or in the availability of these objects to them. Again, since the "true" sizes of these three objects were somewhat indeterminate, it was felt that personal factors might enter to

[1] I wish to thank the staffs of the Veterans Administration Hospitals at Palo Alto, California, Fort Custer, Michigan, and Ann Arbor, Michigan, for their cooperation.

TABLE 1
"SYMBOLIC ENHANCEMENT" SCORES

Group	Mean	*SD*
Control	1.03	4.46
Paranoid	3.30	4.27
Nonparanoid	3.50	5.88

TABLE 2
DIFFERENCES IN "SYMBOLIC ENHANCEMENT" SCORES

Group	*t*	*p**
Control vs. paranoid	1.92	.03
Control vs. nonparanoid	1.80	.04
Control vs. total schizophrenics	2.24	.01

* Based on one-tailed test of hypothesis that patient groups will score higher than controls.

a greater extent than for choices based on less variable items.

Hypothesis. It was hypothesized that symbolic value exerts more influence in the perceptual judgments of schizophrenics than is true for normals. The prediction was that the schizophrenics as compared to controls would tend to judge symbolic objects as larger in relation to their judgment of the neutral object. The hypothesis, it should be noted, deals with the discrepancy between symbolic and neutral object choices, rather than with the choices themselves; and it relates only to psychotic and control populations without regard for the paranoid-nonparanoid distinction.

Results and discussion. A score was obtained for each subject by summing the differences between each of the symbolic object choices (i.e., the cards chosen to represent the sizes of an average scoop of ice cream and the thickness of an average cigar) and the neutral choice (the card chosen to represent the size of an average overcoat button). The higher this score, the greater is the subject's enhancement of the symbolic objects with respect to his judgment of the neutral object. Table 1 presents means and standard deviations for the three groups of subjects.

Table 2 gives *t*'s and probability levels based on the hypothesis that the patients would show greater enhancement of the symbolic objects as compared to the neutral object than would the controls. The results are as predicted: The patients as compared to the controls overjudge the symbolic objects in comparison to the neutral object ($p = .01$ for the total group of schizophrenics).

The differences in scores between patients and controls are clearly significant. They are small, however, and there is considerable overlap among groups. Since the objects are familiar, since there are no ambiguities with respect to distance, and since the patients are relatively well oriented, one can hardly expect differences of great magnitude. It is unlikely that an average scoop of ice cream would be judged to be the size of a pea or the size of a basketball. It is important to note that the groups do not differ significantly with respect to the average sizes chosen for either the overcoat button, scoop of ice cream, or cigar. Only when the discrepancy scores between the symbolic objects and the neutral object are compared do significant differences appear. It would seem that, relative rather than absolute discrepancies differentiate the groups in this rather simple situation.

The experiment was repeated at another hospital in a different location. The subjects were 31 male schizophrenic patients with 43 mostly nonprofessional male hospital personnel as controls. This time five objects were judged. The difference between symbolic objects (scoop of ice cream, cigar, and milk bottle cap) and neutral objects (overcoat button and spool of thread) is again greater for the schizophrenics, yielding a t of 1.69, significant at the .04 level on a one-tailed test. Here too, the scoop of ice cream and the cigar diameter tend to be judged larger and the overcoat button tends to be judged smaller by the schizophrenics than by the controls, although none of these differences is significant. The schizophrenics, however, choose a significantly larger circle to represent an average, milk bottle cap than do the controls ($t = 4.49$, $p < .001$). Thus, replication bears out the original findings.

The results of these experiments lead to the belief that object distortion in schizophrenics is related not only to ego modifications, but to the specific content of the stimulus material. One would expect such distortions to increase with loss of reality values and to vary as a function of the relative cathexes of objects or their psychic representatives.

In a sense, these experiments are little more

than a demonstration of the interaction between specific stimulus factors and ego processes. While the results are consistent, their magnitude indicates that the processes are not defined by a dichotomy between "schizophrenic" and "normal," although they are related to whatever factors are involved in this differentiation. Whether the more genotypic variable is type of ego-control (2, 3), or control by constriction vs. flexible control (7), or some other dimension, remains a question. Furthermore, the demands of the task are more or less on accuracy. The overlap between schizophrenics and normals indicates that with some exceptions this demand can be met by patients diagnosed as schizophrenics. From this study we can say nothing about schizophrenic perceptual or thought processes when such demands are not made.

STUDY 2: CONCEPTUAL STABILITY

Procedure. After the subject had chosen the circle representing his guess as to the size of an average overcoat button, he was asked to choose the circle which would represent his *largest* guess as to the size of an average overcoat button; next, he was asked to choose the circle which would represent his *smallest* guess as to the size of an average overcoat button. The same procedure was followed for the average scoop of ice cream and for the average cigar.

Hypotheses. This cognitive situation is one in which the cues are imperfect, since the circle representing the average cannot be known absolutely. Indeed, the instructions, by asking for the largest and the smallest guesses as to the average, imply that there is ambiguity involved in a guess. The problem, therefore, has to do with the differences among the three groups in dealing with the concept of a guess as to an average. On the basis of theoretical and clinical considerations, it was hypothesized that paranoid schizophrenic patients in their handling of the concept demonstrate rigidity as compared to the controls in that their largest and smallest guesses would deviate less from their original guesses than would be true for the controls. On the other hand, excessive fluidity was hypothesized for the nonparanoid schizophrenics in the sense that their largest and smallest guesses would deviate more from their original guesses than would be true for the controls.

Results and discussion. Scores were obtained for each individual by summing the differences between his largest and smallest guesses for each of the three objects. The scores arrived at in this fashion present rather skewed distributions, particularly for the paranoid group, where they mass toward the zero end of the scale. The nonparanoid distribution seems similar to that of the controls, but with a wider spread of scores. Medians for the 30 controls, 30 paranoid, and 30 nonparanoid schizophrenics are 6.88, 4.00, and 6.30, respectively; semi-interquartile ranges for the three groups are 1.04, 2.38 and 3.50, respectively.

Comparisons were made by means of chi square. When the total distribution is divided into essentially equal halves, the scores of the paranoids are significantly lower than those for the controls ($\chi^2 = 9.64$, 1 *df*, $p < .01$). The nonparanoids, however, do not differ from the controls. Consequently, the hypothesis that the paranoid schizophrenics as compared to the controls show smaller deviations from their original selections is confirmed, while the hypothesis that the nonparanoid schizophrenics show larger deviations is unsupported.

In a size constancy experiment with the same subjects (12), the hypothesis that the paranoids would show higher constancy scores than the controls was verified. The prediction of lower constancy for the nonparanoid schizophrenics failed to be demonstrated. It is of interest here to examine the relationship between size constancy performance and stability in dealing with a concept. In both situations, the paranoids react to the ambiguity by inflexibility. In the perceptual task they achieve, as a group, the highest constancy scores; in the conceptual situation their deviations from their original choices are, as a group, the smallest. The scores of the nonparanoids are intermediate in both situations, and in neither situation are they as predicted. The controls, as a group, show the lowest constancy scores and the greatest fluidity on the conceptual task. The within-group correlations between constancy scores and scores on the concept task are in the expected negative direction, but they are not significant.

These results with both perceptual and

cognitive measures of stability tend to confirm theoretical, clinical, and experimental (6, 8, 11, 14) suggestions that paranoid defenses are associated with the projection of an artificial stability onto the environment. Whether the regressive phenomena predicted for the nonparanoid schizophrenics, i.e., loss of object constancy and excessive conceptual fluidity, are to be found in less well-oriented and more acutely disturbed subjects than those of this study remains to be demonstrated. Difficulties in communication and motivation with such subjects make experimental arrangements and interpretation of data difficult.

SUMMARY

The subjects involved in the two studies presented here were 30 college student controls, 30 paranoid schizophrenics, and 30 nonparanoid schizophrenics. The first study dealt with the hypothesis that symbolic value has greater influence on size judgments made by schizophrenic patients than on those made by controls. The hypothesis was confirmed, and a repetition of the experiment with other patients and controls yielded similar results. The second study dealt with the problem of stability of conceptual judgments. It was hypothesized that (*a*) paranoid schizophrenics show greater constriction in their judgments as to the range of a concept than do controls, and (*b*) nonparanoid schizophrenics show greater fluidity in their judgments as to the range of a concept than do controls. The first hypothesis was verified; the second failed to be demonstrated. The relationships between conceptual stability performance and performance in a traditional size constancy experiment were examined.

REFERENCES

1. ANSBACHER, H. Perception of number as affected by the monetary value of the objects: a critical study of the method used in the extended constancy phenomenon. *Arch. Psychol.*, 1937, **30**, No. 215.
2. BLOCK, J., & BLOCK, JEANNE. An investigation of the relationships between intolerance of ambiguity and ethnocentrism. *J. Pers.*, 1951, **19**, 303–311.
3. BLOCK, JEANNE, & BLOCK, J. An interpersonal experiment on reactions to authority. *Hum. Relat.*, 1952, **5**, 91–98.
4. BRUNER, J. S., & GOODMAN, C. C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33–44.
5. BRUNSWIK, E. *Wahrnehmung und Gegenstandswelt—Grundlegung einer Psychologie vom Gegenstand her.* Leipzig: Deuticke, 1934.
6. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108–143.
7. KLEIN, G. S. Need and regulation. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1954.* Lincoln, Nebr.: Univer. of Nebraska Press, 1954, Pp. 224–274.
8. KLEIN, G. S., & SCHLESINGER, H. Where is the perceiver in perceptual theory? *J. Pers.*, 1949, **18**, 32–47.
9. KLEIN, G. S., SCHLESINGER, H. J., & MEISTER, D. E. The effects of personal values on perception: an experimental critique. *Psychol. Rev.*, 1951, **58**, 96–112.
10. KRECH, D., & CRUTCHFIELD, R. S. *Theory and problems of social psychology.* New York: McGraw-Hill, 1948.
11. MCREYNOLDS, P. M. Perception of Rorschach concepts as related to personality deviations. *J. abnorm. soc. Psychol.*, 1951, **46**, 131–141.
12. RAUSH, H. L. Perceptual constancy in schizophrenia: size constancy. *J. Pers.*, 1952, **21**, 176–187.
13. SINGER, J. L. Personal and environmental determinants of perception in a size constancy experiment. *J. exp. Psychol.*, 1952, **43**, 420–427.
14. VOTH, A. C. An experimental study of mental patients through the autokinetic phenomenon. *Amer. J. Psychiat.*, 1947, **103**, 793–810.

Received April 7, 1955.

SOME FACTORS AFFECTING AN INDIVIDUAL'S ESTIMATE OF HIS PROBABLE SUCCESS IN A GROUP SITUATION[1]

HAROLD B. GERARD

University of Buffalo

THIS study is concerned with some factors which may affect an individual's subjective likelihood estimate in assessing the success of a prospective group endeavor.[2] The choice of estimate the individual makes will be determined, in part, by characteristics of the information available to him and, in part, by expectations which he has brought with him into the judgmental situation. Previous experience of successful goal attainment in similar situations can be expected to produce favorable expectations, whereas previous failure can be expected to produce unfavorable expectations. In a situation, then, where an individual is assessing group action as a path to some goal the hypothesis may be posed that *the more successful an individual's past group experience, the more favorable is his subjective likelihood estimate of subsequent success via group action.*

The individual's set can be conceived of as the probability distribution of the possible evaluative categories to be fitted to the mediated information. He may experience considerable difficulty in evaluating a particular path of action. The mediated information may be completely strange and unfamiliar to him. In that case all possible evaluative categories are equally likely. More normally, however, judgmental difficulty arises in situations which are only partially equivocal. Some small number or limited range of possible evaluative categories might be appropriate. If the situation concerns group action as a path toward some goal it may be hypothesized that *the more equivocal the information available to* S *concerning an alternative course of action open to him, the less certain he is of his subjective likelihood estimate of success via that course of action.*

It has been suggested that prior experience will induce certain expectations. Now if there exists an environmental basis for multiple evaluation, a relatively greater number of evaluative categories may seem appropriate to *S* than if the mediated information is less equivocal. More evaluative categories which have been influenced by experience will tend to be included among the greater number which are compatible with the information. The literature contains many examples of studies dealing with the relationship between expectations, equivocality of the stimulus and evaluation (e.g., 1, 5). The general conclusion from these studies is that, confronted by equivocality, the individual tends to evaluate his world in a manner which is consistent with his expectations. The hypothesis is offered here that *the greater the equivocality of the situation being judged, the greater is the influence of set resulting from past experience in similar situations.*

METHOD

Subjects

Eighty-four United States Air Force enlisted men served as subjects in the experiment. They were chosen so as to be above the general level of intelligence in the Air Force.

Procedure

Six *S*s were scheduled to appear in the laboratory and were seated in a room for the purpose of experimental briefing. The instructions were as follows:

> Each of you will be assigned to work with two other people on a problem involving manipulating objects in space. We haven't as yet made the group assignments, but we will do that very shortly and then inform you as to what they are. The ability involved in working on this problem has been investigated very thoroughly by psychologists. They call it spatial visualization. They have discovered something very remarkable, namely that just about half of the population at large has the ability whereas the other half has not. It's an either-or-thing; you have it or you haven't got it. Now, if you happen to get into a group where all

[1] This research was conducted under a grant from the Office of Naval Research (Contract No. NONR 283 10). The author is indebted to Dr. Morton Deutsch for helpful criticisms, to Mr. James Loomis for help with the analysis, and to Miss Ruby Weinberg for help in running the experiment.

[2] In this discussion we are concerned not with whether the individual is correct in his expectations, but with what the expectations really are. He may be completely wrong in his estimate of the situation. An erroneous estimate may have future consequences for him and for others, but is another problem.

three members have this spatial ability, your chances for succeeding at the puzzle are excellent. However, if your group is composed of three people who haven't got the ability, your chances for success are indeed very slim. In the past we have found that only half of these groups succeeded in completing the puzzle in the alloted time. This appears to reflect the possession or lack of spatial visualization. I can't tell you yet just what the puzzle is, only that it involves this ability.

Psychologists have also discovered that if an individual has spatial visualization, he knows he has it. This, in and of itself, is also a remarkable discovery. You will get a chance to learn who among you has the ability and also you will be able to tell the others whether or not you have it. You will each be assigned to a separate cubicle in which there is a pencil and a pad of note paper. When you get into the cubicle I want you to write some notes to the others telling them whether or not you have spatial visualization and how you know.

The Experimental Conditions

Success-failure treatment. The method employed to induce prior experience of success and failure utilized an actual group experience.[3] The *S*s in an experiment (3) being conducted concurrently by a colleague were exposed to two successive group experiences in which they were told that they either succeeded or failed on both tasks. This announcement was arbitrary and not dependent upon the actual group scores. The *S*s then immediately took part in the present experiment.

Equivocality of the pregroup situation. The six *S*s were ushered into adjoining cubicles and were told to write some notes indicating whether or not they had spatial visualization and how they knew. In the nonequivocal treatment *S* was asked to write two notes, each of which would be delivered by *E* to the two people with whom he had been assigned to work. Each *S*, therefore, would in turn receive two notes from his prospective coworkers.

For the equivocal pregroup situation each *S* was asked to write four notes indicating whether or not he possessed the ability and how he knew. These four notes were to be delivered to four of the five other *S*s. They were instructed that two of these four people would be assigned to work with the sender of the notes. Each *S*, therefore, would receive four notes from four different *S*s, only two of whom would be assigned to work with him. The decision as to the actual composition of the two three-person groups, they were told, would be made shortly thereafter. All notes written by the *S*s were intercepted and false notes were substituted and delivered. The same set of false notes were used throughout the experiment:

1. I guess I'm one of those people who has the ability he was talking about. That has always been one of my strong points since as far back as I can remember. I've always enjoyed working on puzzles of that kind and always did well on them.

2. I know I'm good at this sort of thing. Whenever I've had to work on such problems, they're always a snap. Some people I know have a lot of difficulty with them, but at least that's one thing I'm really good at.

3. You've really got an eight-ball if you get me in your group. Puzzles like that throw me. I must be one of the have-nots when it comes to that space ability he was talking about.

4. Since we are going to work together I ought to tell you that I've always had trouble with puzzles. Whenever I have to do one I get mixed up and can't fit the pieces together.

Notes 1 and 2 indicate possession of the ability, whereas notes 3 and 4 indicate a lack of it. Each *S* in the equivocal situation received all four notes and anticipated that two of the four people who wrote them would be working with him. He was unable to tell, however, whether the two chosen for his group would both have the ability, both lack it, or only one would have it. The probability that an *S* in the equivocal treatment would have one associate possessing and the other lacking the ability was 0.5, that he would have both with the ability was 0.25, and that he would have both lacking the ability was 0.25.

Each *S* in the nonequivocal situation received two of the four false notes, one of the notes indicating possession of the ability and the other indicating a lack of it. The four such possible combinations of notes (1 and 3, 1 and 4, 2 and 3, 2 and 4) were distributed randomly so that the two treatments would be comparable in the analysis. Thus, the expected value of the quality of the group's joint skills to perform the task is the same in both the equivocal and nonequivocal conditions.

The Measures

The *S*s were asked to read the notes very carefully. The *E* entered the cubicle and announced that he was going to try to determine what *S* thought his chances were for success on the puzzle.

One of the methods used to determine *S*'s subjective likelihood estimate of success involved the presentation of white unlined 5 x 8 cards with the odds for success boldly printed on them in black india ink. The odds printed on the cards were as follows: 10-to-1, 5-to-1, 2-to-1, 1-to-1, 1-to-2, 1-to-5, 1-to-10. The instructions proceeded: "I have seven cards in this deck. Each has different odds for success printed on it. After reading the notes you received you have some idea as to what your group's chances are for succeeding at the puzzle. I am going to present these cards to you two at a time. When I show you a pair of cards, I want you to tell me which odds better reflects what you think your group's chances are for success. Odds of 10-to-1 indicates a very optimistic judgment whereas odds of 1-to-10 indicates pessimism."

The *E* did not proceed with the administration of the cards until he was certain that *S* understood the procedure and was aware of what the odds meant. All 21 possible pairs (7C2) were presented and *S*'s choices were recorded. After this procedure, a questionnaire (subsequently to be discussed) intended to provide

[3] Additional subjects were run under conditions of vicarious success and failure experience. The *S*'s read stories in which the main character experienced success or failure via group action. The rationale for the use of these stories was that *S* would presumably identify with the main character and share his experience. This method of induction proved unsuccessful.

another measure of *S*'s subjective likelihood estimate was administered. When the experiment was over, *E* described the purpose of the experiment and revealed the nature of the deceptive devices. The *S*s appeared to react good-humoredly and promised not to disclose the deception to others.

Summary of the Design

The 84 subjects used in this experiment were exposed to the same general experimental setting. Prior experience of success and failure in a group were induced by providing an actual experience of success or failure. Half of the *S*'s were exposed to an equivocal situation, whereas the other half were exposed to a nonequivocal situation. The experimental treatments and the *N*'s are indicated below.

	Success	Failure
Equivocal	24	18
Nonequivocal	21	21

RESULTS

The paired comparison data were used in testing the first hypothesis, that the more successful an individual's past experience with group action, the more favorable his subjective likelihood estimate of future successful group action. The individual's rank placement of the most favorable stimulus card (10-to-1) was taken as an indication of his subjective likelihood estimate, since most *S*s appeared to pay especial attention to it during administration of the cards. Since each stimulus card was paired with each of the other six cards, a rank order is easily generated from any total set of comparisons.

The data for each subject were organized in the form of a matrix. A zero was entered in every cell of the diagonal; +1 was entered in any cell if the row stimulus card was preferred to the column stimulus card; −1 was entered if the column stimulus card was preferred to the row stimulus card. The entries above and below the diagonal are images of each other with opposite signs, reflected from the diagonal. Ordering the row totals generates the individual's rank order of preference. A score of +6 for a card indicates that these odds were preferred to all others. A score of −6 for a card indicates that all other odds were preferred to it.

The rank placement of the 10-to-1 stimulus card was averaged for each experimental condition. These averages are presented in Table 1. The table indicates a tendency for successful past experience with group action to result in a more favorable estimate of future success than does past experience of failure ($p < .001$ by t), thus supporting the hypothesis.[4]

TABLE 1

AVERAGE RANK PLACEMENT OF THE MOST FAVORABLE ODDS (10 TO 1) FOR SUCCEEDING

	Success	Failure
Equivocal	3.67	3.97
Nonequivocal	3.09	5.12

In the final questionnaire *S* was confronted with a series of imaginary bets. In each of the bets *S* had $10.00 to wager. One of the bets was for "even money," that is, ten dollars to be "covered" by *S*'s ten. On the optimistic side, bets were available ranging from 10-to-1 to 10-to-9 by single steps, whereas the bets provided on the pessimistic side ranged from 10-to-20 to 10-to-100 in steps of 10. This arrangement produces some confounding asymmetry but as a crude corroborative measure it appeared to be adequate. The *S* was asked to indicate, for each bet, whether it was favorable or unfavorable to him or whether he was uncertain as to its favorableness. For each *S* we determined whether there were more favorable or unfavorable bets. A chi-square value was computed for the preponderance of either favorable or unfavorable bets for the success and failure treatments. Those few *S*s who indicated an equal number of favorable and unfavorable bets were arbitrarily included in the unfavorable category. There were a greater number of *S*s indicating more favorable than unfavorable bets only in the success treatment ($p < .001$). These data lend additional support to the first hypothesis.

Information from the preceding data can be used to test the second hypothesis, which concerns an individual's confidence in his

[4] Another measure of favorableness of set using the entire rank order generated by these paired comparison data involved the assignment of weights to the stimuli. The stimulus weight was multiplied by the row total for that stimulus. The product of weight and row total was then summed for the seven stimuli. This measure, although it utilizes considerably more information, makes some untenable assumptions about favorableness of set as an attribute. For this reason these figures are not presented. Still another method was based on the mean of weighted first choices within each experimental treatment. There is close agreement among all three measures.

TABLE 2

PROPORTION OF SUBJECTS REVEALING INTRANSITIVITIES AND/OR ORDERING ERRORS IN PAIRED COMPARISON JUDGMENTS

	Success	Failure
Equivocal	.46	.67
Nonequivocal	.14	.19

subjective likelihood estimate. It appears reasonable to assume that the more certain an individual is of his subjective likelihood estimate for success, the less will be his tendency to waver, be indecisive, or change his mind during the presentation of the paired comparison stimulus cards. One type of inconsistency which may occur in paired comparison data is intransitivity of choice. In the paired comparison data, intransitivity of choice may be taken as an indication of uncertainty regarding the future success of the group. Kendall (4) describes a method of counting the number of intransitive triplets that occur in a complete set of paired comparisons. The number of such errors committed by each *S* were counted.

If it is assumed that the stimuli possess a natural order from extreme optimism to extreme pessimism, another type of error is possible. An *S* may be completely transitive in making his judgments but he may have ordered the stimuli in a manner which is inconsistent with the assumed natural order. Whether or not an intransitive triplet occurs is completely independent of any ordering, whereas an ordering error is based upon the assumption of a natural order.[5] Only five *S*s reordered the stimuli.

The proportion of inconsistent matrices (those containing intransitive triads and/or ordering errors) which occurred in the equivocal and nonequivocal treatments are presented in Table 2. A chi-square test indicates that a greater proportion of inconsistent choice patterns occurred in the equivocal treatment ($p < .001$). This evidence supports the second hypothesis.

[5] Since a description of the method for counting intransitivities and ordering errors would unduly lengthen the present report it has been omitted. An article describing the technique, including proofs for the formulae, is in preparation.

In the questionnaire item confronting *S* with a series of bets, data bearing on the same hypothesis were available. There was a tendency for *S*s in the equivocal treatment to be more uncertain as to the favorableness of the bets than were *S*s in the nonequivocal treatment ($p < .01$ chi-square). The *S* also indicated, by circling a point on a seven-point a priori scale, how certain he was of his estimate of the group's future performance. There was a tendency for greater equivocality to produce more uncertainty ($p < .07$ by t). These findings lend additional support to the hypothesis.

In the equivocal pregroup situation, it has been also hypothesized that *S*'s judgments are more susceptible to influence resulting from prior group experiences. Presumably these influences varied randomly between the equivocal and nonequivocal treatments. In an equivocal judgmental situation where a greater range of possible evaluative categories exists, there is greater freedom for the individual to bring what he sees into line with what he expects to see. A greater range of favorableness of set is therefore predicted in the equivocal than in the nonequivocal treatment. The variance of the data presented in Table 1 was therfore computed. A simple F test, comparing the variances in the equivocal and nonequivocal treatments, yielded a significant difference in the predicted direction ($p < .05$). This finding lends support to the third hypothesis.

This hypothesis predicts that set, as systematically varied by the success-failure inductions, tends to influence *S*'s subjective likelihood estimate to a greater extent in the equivocal than in the nonequivocal treatment. Thus *S* should be more optimistic in the equivocal-success treatment than in the nonequivocal-success treatment, with parallel results for failure. An examination of the data in Table 1 indicates that this prediction is unsupported. In fact, there is a striking reversal; favorableness of set exerts a greater influence in the nonequivocal treatment. A plausible interpretation to account for this apparently paradoxical finding can be suggested which involves the relative strength of *S*'s judgmental referents. It has been shown that *S* tended to be confused in the face of uncertainty regarding the composition of the group (see Table 2),

and that this confusion tended to be accompanied by greater random fluctuations of subjective likelihood estimates. This finding was interpreted, in accordance with the third hypothesis, as resulting from expectations which the *S*s had brought to the experimental situation. When confronted with a specific but confusing situation, more generalized expectations may tend to have a stronger influence upon evaluation than do expectations derived from immediately prior experience. The relatively stronger influence of these extra-experimental expectations may have resulted in a regression to some "typical" evaluations accounting for the unanticipated results. In the nonequivocal situation there was less confusion and hence possibly a greater reliance upon expectations derived from a similar immediately prior experience. Additional experimental evidence, potentially obtainable, is required in order to confirm this suggested interpretation. The relatively weak and specific nature of the induced sets is a common shortcoming in laboratory experimentation using human subjects.

There is at least one other possible interpretation of this result. Although none of the tasks performed in the prior experiment directly involved "spatial visualization," experience of success or failure in them may have induced differential, personal feelings of general skill adequacy. The questionnaire administered after the paired comparisons provides data, however, indicating that prior success or failure did not induce such effects. The *S* was asked, "Are you good at putting jig-saw puzzles together?" This question, which could be answered "yes" or "no" showed no difference between the success and failure treatments. The ensuing question asked *S* to rate his performance on this kind of problem on a 15-point scale, ranging from "extremely well" to "extremely poor." Again no difference occurred between the success and failure treatments. An analysis of the content of the notes written by the *S*s concerning their "spatial visualization" ability also yielded no difference between the success and failure treatments.

SUMMARY AND CONCLUSIONS

Some factors influencing an individual's view of the likelihood of successful group action were studied in a laboratory experiment. The *S* was given a prior experience of either success or failure via group action. The *S* also faced either equivocal or nonequivocal information with regard to the composition of his group with respect to skills required by the experimental task. In the equivocal treatment he knew that he was to be in a group with two of four possible people. In the nonequivocal treatment he was sure of the group's composition. In either case the expected value of the quality of the group's joint skills to perform the task was the same, i.e., the expectation was that one of the other two would have the required skill and the other would not have it.

The data supported the following two hypotheses: (*a*) the more successful an individual's past group experience, the more favorable is his subjective likelihood estimate of subsequent success via group action, and (*b*) the more equivocal the information possessed by the individual concerning an alternative course of action open to him, the less certain is he of his subjective likelihood estimate of success via that course of action. A third hypothesis was given partial support: namely, that the greater the equivocality of the situation being judged, the greater is the influence of set resulting from past experience in similar situations.

REFERENCES

1. BRUNER, J. S., & GOODMAN, C. C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33–44.
2. CARTWRIGHT, D. Decision time in relation to the differentiation of the phenomenal field. *Psychol. Rev.*, 1941, **48**, 425–442.
3. DEUTSCH, M. The effects of different probabilities of success upon the cooperative process under varying conditions, in press.
4. KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.
5. LEVINE, R., CHEIN I., & MURPHY G. The relation of the intensity of a need to the amount of perceptual distortion. *J. Psychol.*, 1942, **13**, 283–293.

Received April 18, 1955.

FUTURE TIME PERSPECTIVE IN SCHIZOPHRENIA[1]

MELVIN WALLACE

Veterans Administration Regional Office, Kansas City, Missouri

In recent years one phase of psychological research has been devoted to attempts to investigate phenomena associated with temporal experience, especially in terms of their relevance to personality functioning. Although the number of investigations in this area has become relatively sizable, many questions remain to be answered before an adequate basis can be provided for any systematic understanding of the complex factors involved. At least in part, such a state of affairs seems to be attributable to a lack of clarity and preciseness in defining those temporal notions which have been selected for study. Concepts such as *time sense* (2, 3, 7), *time orientation* (1, 8, 20), *time perspective* (10, 11), and *time perception* (4, 5) are employed interchangeably, often in the same investigation, or are utilized in such a manner that no clear idea of the intended meaning is given, either in conceptual or operational terms.

To eliminate this lack of rigor in the definition of temporal concepts, the study to be reported focuses upon one segment of the temporal dimension, the future. To restrict further the range of temporal phenomena to be examined, attention is devoted to only one aspect of handling conceptions of the future, termed *future time perspective*. Partially derived from views presented by Lewin (17) and Frank (12), this concept is defined in the present context as the timing and ordering of personalized future events. In a further attempt to delineate more precisely those aspects of *future time perspective* to be investigated, two heretofore unspecified, although often implied concepts, *extension* and *coherence*, are also introduced. The former is defined as the length of the future time span which is conceptualized. *Coherence* is utilized to refer to the degree of organization of the events in the future time span.

An examination of the pertinent literature indicated that direct references to the relationship between distortions in handling conceptions of the future and psychopathological conditions are almost completely lacking. The reports of Israeli (14) and Escalona (9), and the theoretical views of Meerloo (18) and Schilder (19) are relevant; but neither these accounts, nor the implicit references to the future found in psychoanalytic (e.g., 21), social psychological (6, 16), and sociological (13) formulations concerning the evolution of temporal notions provided a suitable theoretical basis from which predictions of a directional nature could be derived. As a result, the following nondirectional[2] hypothesis was proposed: If *future time perspective* is significantly affected by mental disturbance, then schizophrenics differ from normals in *extension* and *coherence*.

METHOD

Subjects. The control group consisted of 34 *S*s chosen from the medical service of a Veterans Administration General Medical and Surgical Hospital. Although the range of medical and surgical problems for which the patients in this group were being treated was quite wide, including such conditions as arm and leg fractures, "trick" knees, nephritis, and hernia, none of these *S*s had any record of psychiatric illness. Their length of hospitalization varied from one week to eight months, with the median at two months.

The experimental group comprised 34 in-patients at a Veterans Administration Neuropsychiatric Hospital. Each patient had an official diagnosis of schizophrenia or schizophrenic reaction with no history of organic involvement. Only those psychotics judged by physicians and ward personnel to be sufficiently cooperative were asked to participate in the study. Because eight

[1] Part of a dissertation presented as partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the Department of Psychology, Michigan State University. The valuable advice and critical comments of Dr. Albert I. Rabin, chairman of the writer's doctoral committee, Dr. Marvin Hyman, and the Psychology Staff of the Detroit, Michigan, VA Mental Hygiene Clinic, and the cooperation of the psychology departments at the Dearborn and Battle Creek VA Hospitals in Michigan, are gratefully acknowledged.

[2] Another consideration also played a role in the decision to utilize nondirectional predictions in the present study. It was reasoned that if the tests of nondirectional hypotheses produced statistically significant results, the opportunity was then provided for examining the data on an a posteriori basis to determine the directionality of the obtained intergroup differences. While such findings obviously could not be considered to be established scientific facts, they could nonetheless be utilized as a basis for making new directional hypotheses which could be investigated in further research.

patients could not or did not wish to respond to the experimental tasks, their performances were excluded from consideration in the total sample.[3]

In order to take into account the factor of length of hospitalization, the sample of schizophrenic patients was divided into two groups of 17 *S*s each. For the first of these groups, the length of time spent in the hospital varied from one week to 23 months, with six months as the median. With regard to the remaining experimental *S*s, however, the duration of hospitalization was found to range from five years and four months to eleven years and one month. For this group, the median period of hospitalization was five years and nine months. Tests of significance with regard to several pertinent variables (occupational level, educational level, and Wechsler vocabulary score) revealed no significant intergroup differences. With regard to age, however, the differences between the two subsamples reached a statistically significant level.

The experimental and control groups were equated for the following variables: (*a*) age; (*b*) vocabulary level (Wechsler, Form I); (*c*) educational level; and (*d*) occupational level. A statistical analysis of the data with regard to these four criteria of selection revealed no statistically significant differences between the two groups.

Procedure. This experiment was composed of three tasks, each of which was administered to every *S*.

Task I consisted of two parts. On the first part, *E* gave the following instructions: "Tell me ten events that refer to things that may happen to you during the rest of your life." Each event given was recorded on a separate card, and its ordinal rank in the series of events was noted. After each response was given, *E* asked *S*: "And how old might you be when that happened?" The age given for each event in the series was recorded on a separate sheet.

The second part of Task I was carried out after the remaining tasks, Task II and Task III, had been completed. The *E* returned to *S* the ten cards on which the events given previously had been written. The following instructions were then given: "Arrange these cards in the order in which they might occur." The sequence of events presented by *S* was recorded.

The following two measures were based upon data obtained on Task I: (*a*) ***extension***: the range of years included between *S*'s actual age and the most distant event given by him, and (*b*) ***coherence***: the correlation between the ranking of events based upon the age of occurrence given in Part 1 with the order of events given in Part 2.

Task II consisted of four story-completion items which were developed from a technique utilized by Barndt (1). The first two stories involved structural temporal situations; the third and fourth items did not contain explicit references to the temporal dimension. The following instructions were given *S*:

Story 1

I want to see what kind of a story you can tell. I'll start one for you, and then let you finish it any way you wish. I'll start it now. "At three o'clock one bright sunny afternoon in May, two men were walking near the edge of town. . . ." Now you start there and finish the story for me.

Story 2

That was fine. Now I'll begin another story which, as before, you may finish any way you want to. Here it is: "Ten o'clock one morning Al met his friend Jerry near the center of town. . . ." Now you start there and finish it for me.

Story 3

That was pretty good. Now here is the start of another story which you may again finish in any way you wish. "Joe is having a cup of coffee in a restaurant. He's thinking of the time to come when. . . ." Now you finish it for me.

Story 4

Here is the last story that I'll give you. I want you to finish it any way you wish, just like you did on the other stories. Here it is: "After awakening, Bill began to think about his future. In general he expected to . . . ." Now you start there and finish it for me.

In each case the story given by *S* was recorded. If no reference to the interval of time involved in the story was included, *E* asked: "How long a time was involved in this story—not in telling it, but in the action described?" A record was made of the response to this question for each part of Task II.

One measure of ***extension*** was derived from each story-completion item on Task II. In each case, the obtained ***extension*** score referred to the length of time given by *S* in describing the duration of the action in each story. Since *S*'s response to each part of Task II might conceivably be expressed in terms of varying time-units (e.g. minutes, months, years), the score for each item was treated separately in subsequent analysis, with no attempt being made to compute a summated or averaged index of performance.

Task III is composed of 15 statements describing common life-events. These items were originally chosen on the basis of logical considerations regarding the likelihood of occurrence of such events in the general population of American males. Later refinements, which led to the technique in its present form, stemmed from exploratory work with normal and schizophrenic subjects.

The *S* was given the following instructions: "I'm going to read to you a series of events which happen to a lot of people. I want you to tell me how old you might be when the event described might happen to you." Before reading each event, *E* said: "How old might you be when. . . ." The *S*'s response to each event was recorded.

Items

1. Your first grandchild is born?
2. You die?
3. You lose interest in sexual activities?
4. You are too old to be physically active?
5. Your son achieves his greatest success in life?

[3] Six of these eight schizophrenic patients were unable to offer future events that might take place in their lives (see Task I below) even though they were quite communicative in other respests. Although not included in the final sample, their performances, it is to be noted, tended to support the hypothesis under test. The remaining two patients who were omitted from the experimental sample did not wish to leave the ward with the experimenter.

6. You retire?
7. Your youngest child leaves home?
8. Your mother dies?
9. You feel that you have reached old age?
10. Your last sexual intercourse occurs?
11. You have reached middle age?
12. Your father dies?
13. You can no longer have children?
14. Your first great-grandchild is born?
15. You can say that you have most of the things you want?

After *S* had answered all of the above items, he was presented with a series of 15 cards upon each of which one of the test items was printed. He was then told:

> Here is a group of cards upon which the events I just asked you about are typed. Arrange these cards in the order that they might happen in your life. Place the event that might happen first, first; the event that might occur next, second, and so on, until you have finished with all the cards.

A record was made of the sequence of cards that was submitted.

The variable of *coherence* was defined as the correlation between the order of events based upon the ages of occurrence associated with them, and the sequence resulting from a forced chronological ordering of the events.

RESULTS

Since the obtained distributions of scores made the assumption of normality questionable, a nonparametric method, the *H* test (15), was employed in analyzing the data. In comparing the performance of the two groups of subjects used in the present study on a particular variable, an *H* score was obtained by ranking all the scores of both samples in a single distribution. On the basis of the sum of ranks for each group, an *H* score was derived, which, when reference was made to tables of normal curve functions, provided the appropriate level of significance for a one-tailed or two-tailed test. For the purposes of the investigation under discussion, a two-tailed test was utilized.

TABLE 1

MEDIANS AND SUMS OF RANKS OF EXTENSION SCORES GIVEN BY 34 NORMALS AND 34 SCHIZOPHRENICS

Task	Normals		Schizophrenics		*H*	*p*
	Mdn	Sum of ranks	Mdn	Sum of ranks		
I	36 yr.	1440.5	12 yr.	905.5	3.28	.001
II						
Story 1	2 hr.	1274	1 hr.	1072	1.24	—
Story 2	2 hr.	1330	1.5 hr.	1016	1.93	.054
Story 3	1.5 wk.	1372	1 hr.	974	2.44	.015
Story 4	4 yr.	1371	.75 yr.	975	2.43	.015

As indicated by the data summarized in Table 1, the results tended to support the hypothesis that schizophrenic patients may be differentiated from normals with regard to the length of the time interval used in conceptualizing future events. It also seems probable that the schizophrenic process limits or curtails the extent of this future time span. The median *extension* score of the control subjects regularly involves a longer time period than the corresponding score given by the schizophrenic patients. The stability of this directional finding, of course, remains to be confirmed by further research.

The degree of relationship among the various measures of *extension* on Task II was found to be .44. No attempt was made to intercorrelate the *extension* scores derived from Task I with those found on Task II since these scores were expressed in terms of a variety of dissimilar time-units which were not amenable to a correlational analysis. Nor was an estimate of reliability obtained which perhaps could have been useful in evaluating the extent to which the negative results found on Stories 1 and 2 of Task II were attributable to the unreliability of the behavior sampled.

In this connection, it is to be pointed out that the bipartite composition of the experimental sample was found to bear no demonstrable influence upon the results obtained. Despite the previously mentioned differences in age and length of hospitalization between the short-term and long-term schizophrenic subsamples, a comparison of their *extension* scores revealed no statistically significant findings. With regard to *extension*, therefore, the two groups were treated as a single homogeneous sample.

In evaluating the experimental results, it must be remembered that the psychotic *S*s used were chosen largely on the basis of their ability to cooperate and to communicate adequately. It is probable, therefore, that they represent relatively *less* severe psychopathological conditions. In consequence, the data are likely to be conservatively biased, making positive findings all the more striking.

Also, on the basis of these considerations, one may attempt to explain the lack of significant intergroup differences found in connection with Stories 1 and 2 of Task II (Table 1). The simplified and highly structured nature

of these two story completion items provide each subject with well-defined situations which lead to only a limited number of future courses of action. Under such conditions, where external demands are easily understood in a context involving no excessive threat, the schizophrenic patient may experience little or no feelings or anxiety. Consequently, he can more efficiently utilize whatever resources he has available, with the result that his performance on the experimental tasks tends to approximate more closely the level of proficiency revealed by the control subjects.

Clinical observations suggest that the schizophrenic is severely threatened when exposed to unfamiliar, unstructured situations which require a specific course of action. At such times, it may be presumed, the schizophrenic individual is beset by an inner turbulence of such an intensity that confusion and disorganization become marked. Since Task I and Stories 3 and 4 of Task II are relatively unstructured, it is possible that they gave rise to circumstances which threatened the experimental *S*s. The resulting emotional disturbance interfered with optimal functioning to such a degree that the magnitudes of the future time span presented by the schizophrenic patients were significantly shorter than those given by the controls. It must be pointed out, however, that before such a rationale can be considered to be genuinely acceptable, further corroborative research is necessary.

From the data presented in Table 2, it may be concluded that the hypothesis concerned with *coherence* has been upheld. These experimental results indicate that the degree of relationship between the actual chronological sequence of a series of future events and a forced chronological ordering of these events is clearly different for schizoprenic patients as compared to the controls. Also of importance is the finding that the median *coherence* score for the latter sample is larger than that found in the schizophrenic group.

The correlation between the measures of *coherence* derived from Task I and those obtained on Task III was found to be .88, thus pointing to a high degree of relationship between the two operational measures of *coherence*.

A supplementary finding. involving a comparison of the two schizophrenic subsamples, points to the fact that these subgroups are homogeneous only with regard to the measure of *coherence* obtained from Task III, where intragroup findings proved to be insignificant (Table 3). On the first experimental procedure, Task I, the obtained *coherence* scores reveal statistically significant intragroup differences. With regard to this measure of *coherence*, therefore, the two subgroups do not comprise a single homogeneous sample, and further analyses were carried out to determine the source of the over-all intergroup differences.

Tables 4 and 5 summarize the results of comparisons of the normal subjects to each schizophrenic subgroup. Table 4 suggests that there are no significant differences in *coherence* scores between the control group and the short-term schizophrenic patients. Table 5, however, indicates that the measure of *coherence* obtained from the controls and the long-term psychotic subjects are markedly dissimilar at better than the .01 level of confidence. It follows, therefore, that with regard to Task I, the confirmation of the *coherence*

TABLE 2

MEDIANS AND SUMS OF RANKS OF COHERENCE SCORES (RHO-CORRELATIONS) OBTAINED FROM 34 NORMALS AND 34 SCHIZOPHRENICS

Task	Normals		Schizophrenics		*H*	*p*
	Mdn	Sum of ranks	Mdn	Sum of ranks		
I	.9257	865.5	.6520	1480.5	3.77	.0002
III	.7766	958	.5464	1388	2.64	.008

TABLE 3

MEDIANS AND SUMS OF RANKS OF COHERENCE SCORES (RHO-CORRELATIONS) OBTAINED FROM 17 SHORT-TERM SCHIZOPHRENICS AND 17 LONG-TERM SCHIZOPHRENICS

Task	Short-term Schizophrenics		Long-term Schizophrenics		*H*	*p*
	Mdn	Sum of ranks	Mdn	Sum of ranks		
I	.8394	202	.4394	393	2.28	.001
III	.5571	262	.4863	333	1.22	—

TABLE 4

MEDIANS AND SUMS OF RANKS OF COHERENCE SCORES (RHO-CORRELATIONS) OBTAINED FROM 34 NORMALS AND 17 SHORT-TERM SCHIZOPHRENICS

Task	Normals		Short-term Schizophrenics		H	p
	Mdn	Sum of ranks	Mdn	Sum of ranks		
I	.9257	806.5	.8394	519.5	.91	—

TABLE 5

A COMPARISON OF MEDIANS AND SUMS OF RANKS OF COHERENCE SCORES (RHO-CORRELATIONS) OBTAINED FROM 34 NORMALS AND 17 LONG-TERM SCHIZOPHRENICS ON ONE EXPERIMENTAL TASK

Task	Normals		Long-term Schizophrenics		H	p
	Mdn	Sum of ranks	Mdn	Sum of ranks		
I	.9257	654	.4394	672	2.65	.008

hypothesis is attributable to the performance of the group of experimental subjects who had been hospitalized the longest.

An over-all comparison of the control and experimental groups suggests that the second aspect of *future time* perspective under examination is also influenced by the schizophrenic process. It appears that, at least on one task, normals are superior to schizophrenics with regard to the ability to order the contents of the future time span logically and meaningfully (insofar as chronological considerations are concerned). A similar finding is to be observed in another experimental task that was utilized, but here the obtained results are to be attributed to the performance of the long-term psychotic subjects rather than the experimental group as a whole.

To meet the criticism that the above experimental findings may reflect merely the effects of institutionalization rather than influences stemming from the operation of the independent variable, two empirical findings may be mentioned: (*a*) Comparisons of the two schizophrenic subgroups revealed no significant differences on any measure of *extension*. With regard to *coherence*, however, significant intragroup differences were found in connection with Task I, but not Task III. On all measures, therefore, except one obtained for *coherence* on Task I, the schizophrenic subsamples are homogeneous. It does not appear that the factor of length of hospitalization produces stable differences in the performances of the psychotic subsamples. (*b*) Since the median period of hospitalization for the short-term psychotic group was six months, while that for the controls was two months, some degree of comparability with regard to this factor may be assumed. Comparisons of these groups revealed statistically significant differences on all measures except for one of *coherence* (Task I). For the most part, then, these findings tend to suggest that factors other than length of hospitalization must be called upon to account for the obtained results. On the basis of these considerations, it appears that the factor of institutionalization per se is not crucial in explaining the results of the present study. Such a conclusion, however, is tentative, and awaits further research in which a more adequate control of this variable is utilized.

SUMMARY AND CONCLUSIONS

One sample of 34 patients, diagnosed as schizophrenic, and a group of 34 patients at a general medical and surgical hospital were given three tasks which involved (*a*) the spontaneous and later a forced ordering of a series of ten future events (Task I), (*b*) four stories in response to verbal instructions which included the beginning of each story (Task II), and (*c*) estimates of the age of occurrence of 15 future events supplied by the examiner which were later placed in order of expected occurrence (Task III). On the basis of these tasks, measures of two aspects of *future time perspective*, *extension*, and *coherence*, were derived.

The results obtained between the experimental and control groups by use of a nonparametric procedure, the H test, indicate that (*a*) on all measures except those found on Stories 1 and 2 of Task II, the scores on both the *extension* and *coherence* variables presented by the schizophrenic group were significantly different from the corresponding scores of the control group, and (*b*) on all tasks, including those which reveal no statistically significant intergroup differences, the

median score of the control group was larger than that of the experimental group.

The obtained results, for the most part, confirmed predictions which were derived from the following general hypothesis: *Future time perspective* is significantly affected by psychopathological disturbance. The lack of significant findings in connection with Stories 1 and 2 of Task II was interpreted on the basis of the fact that these two experimental procedures, in contrast to Stories 3 and 4, were highly structured and presumably relatively free of threat, thus allowing the particular schizophrenic subjects under investigation to function optimally.

On the basis of the present study, it may be suggested that *future time perspective*, in its *extension* and *coherence* aspects, is influenced by the schizophrenic process to such an extent that both the length of the future time span and the organization of its contents are significantly reduced for a sample of schizophrenic patients as compared to a group of normal controls. Additional research is needed, however, before such a finding can be considered to have been conclusively demonstrated.

REFERENCES

1. BARNDT, R. *Time orientation in delinquents.* Unpublished master's dissertation, Michigan State College, 1953.
2. BONAPARTE, MARIE Time and the unconscious. *Int. J. Psychoanal.*, 1940, **21,** 427–468.
3. COHEEN, J. L. Disturbances in time discrimination in organic brain disease. *J. nerv. ment. Dis.*, 1950, **112,** 121–129.
4. COOPER, L. F. Time distortion in hypnosis. *J. Psychol.*, 1952, **34,** 247–284.
5. COOPER, L. F., & TUTHILL, C. E. Time distortion in hypnosis and motor learning. *J. Psychol.*, 1953, **36,** 67–76.
6. DAVIS, A. Socialization and adolescent personality. *Nat. Soc. Stud. Educ., 43rd Yearbook, Part 1.* Chicago: University of Chicago Press, 1944. Pp. 198–216.
7. DOOLEY, L. The concept of time in defence of ego integrity. *Psychiatry*, 1941, **4,** 13–23.
8. EISSLER, K. R. Time experience and the mechanism of isolation. *Psychoanal. Rev.*, 1952, **39,** 1–22.
9. ESCALONA, SYBILLE K. The effect of success and failure upon level of aspiration and behavior in manic-depressive psychosis. In K. Lewin, (Ed.), *Studies in topological and vector psychology.* I. Univer. of Iowa Studies. Studies in Child Welfare, 1940, **16,** 199–303.
10. ESON, M. E. *An analysis of time perspectives at five age levels.* Unpublished doctor's dissertation, Univer. of Chicago, 1951.
11. FINK, H. *The relationship of time perspective to age, institutionalization, and activity.* Unpublished doctor's dissertation, Michigan State Col., 1953.
12. FRANK, L. K. Time perspectives. *J. soc. Phil.*, 1939, **4,** 293–312.
13. HULETT, J. E. The person's time perspective and the social role. *Soc. Forces*, 1944, **23,** 155–159.
14. ISRAELI, N. *Abnormal personality and time.* New York: Science Press Printing Co., 1936.
15. KRUSKAL, W. H., & WALLIS, W. A. Use of ranks in one-criterion variance analysis. *J. Amer. Stat. Ass.*, 1952, **47,** 583–621.
16. LESHAN, L. L. Time orientation and social class. *J. abnorm. soc. Psychol.*, 1952, **47,** 589–592.
17. LEWIN, K. *Field theory and social science.* New York: Harper, 1951.
18. MEERLOO, A. M. Father time. *Psychiat. Quart.*, 1948, **22,** 587–608.
19. SCHILDER, P. Psychopathology of time. *J. nerv. ment. Dis.*, 1936, **83,** 530–546.
20. SCHNEIDER, D. E. Time-space and the growth of the sense of reality. *Psychoanal. Rev.*, 1948, **35,** 229–251.
21. YATES, S. Some aspects of time difficulties and their relation to music. *Int. J. Psychoanal.*, 1938, **16,** 341–354.

Received April 20, 1955.

MOTHER-CONCEPTS AND JUDGMENTS OF YOUNG WOMEN'S FACES[1]

PAUL F. SECORD[2] AND SIDNEY M. JOURARD
Emory University

CONTEMPORARY thinking about social perception suggests that certain personal factors, e.g., values, needs, and cognitive structures, will influence an individual's perceptual-cognitive behavior. The present inquiry involves a study of: (*a*) the relationship between a judge's conception of his mother's personality and the judgments that he makes of photographs of women's faces; and (*b*) the interaction, among the judges, of their mother-concepts with certain aspects of the photographs.

The view that one's mother-concept is related to his perception of women derives from two lines of thought, not necessarily in opposition to one another. The first is represented by Sullivan's (4, p. 101) concept of "parataxic" distortion, which refers to a process of reacting to people as if they represented "significant others" from one's past. The second is the belief that individuals develop, through total life experiences, certain ways of thinking about classes of people. For example, an individual may regard "people-in-general" as hostile, selfish, undependable, etc. Or, he may regard Negroes in particular as stupid, lazy, superstitious, etc. His conceptions of women as a class, then, might lead an individual to rate both his mother and other women in a somewhat similar manner.

From these considerations, the following more specific hypotheses are proposed:

1. *Young males asked to rate the personality characteristics of their mothers will exhibit a high degree of consensus; that is, they will rate in the direction of a cultural stereotype of mother* (which will henceforth be referred to as the "modal mother"). It is not expected that consensus will be complete. Meaningful differences in mother-ratings will be found among the various judges. The profile of mother-ratings characteristic of a particular judge is termed his *mother-concept*. Both the consensus and the differences among judges may reflect concepts of the mother that ar veridical, autistic, or partly both.

2. *Each judge will tend to rate young women unknown to him in the direction of his mother-concept*. More precisely, on a particular trait, his rating of a woman's face is expected to differ from the mean rating of the other judges in the same direction that his mother-rating deviates from the modal mother. Thus, a judge who rates his mother as more *sincere* than do other judges will also rate women higher on the trait *sincere*. This tendency to "distort" ratings of women's photographs in the direction of the mother-ratings is termed *attribution*. The degree to which judges exhibit attribution is thought to depend upon two characteristics of the young women being perceived. Relationships of these characteristics to attribution are stated as corollaries:

2a. *The degree of attribution varies directly with the extent to which a photograph has physiognomic characteristics which "permit" the attribution of mother-characteristics*. In a previous study (3) marked consensus among judges appeared in rating certain photographs high in such traits as *kind*, *sincere*, and *good-mother type*, and also in judging other photographs as lacking in such characteristics. Traits of this kind had been found to intercorrelate highly, and the total correlation cluster was labeled *moral character*. Traits comprising this cluster are those which approximately fit the role ascribed to mothers in our culture; therefore one would expect photographs rated high in moral character to permit attribution of mother-traits, and photographs low or negative in moral character to prevent attribution.

2b. *The degree of attribution varies directly with the ambiguity of the photograph*. Ambiguity is defined as the extent to which a group of judges lacks consensus in rating personality traits from a photograph. An ambiguous or "poorly structured" photograph is thought of as lacking in cues upon which clear judgments might be based, and therefore a photograph which might permit considerable attribution.

A final comment concerning the degree of attribution anticipated should be made. Judgments of photographs are undoubtedly determined by many variables other than attribution; consequently, although a signifi-

[1] This is the seventh paper in a series entitled "Personalities in Faces."

[2] Now at the University of Nevada.

cant amount of attribution is anticipated, it is not expected to be large.

METHOD

Criterion of attribution. As stated in Hypothesis 2, each judge engages in attribution to the extent that, when compared with other judges, he rates photographs of women in the direction of his mother-concept—that is, the extent to which he is above the group mean in rating photographs of women if he is above the mean in rating his mother on a given trait. Since a Pearsonian *r* between the two ratings by the 59 judges is a measure of the extent to which such correspondence in deviations occurs, this familiar statistic was used to measure the degree of attribution for each trait.

Selection of photographs. Four groups of photographs were chosen from 24 photographs of women aged 18–30 years, according to certain personality impressions yielded by them in a previous study (3). Two of these groups represented extremes on a cluster of rating scale traits labelled *moral character*. To select these photographs, composite scores were obtained on the following traits: knows all the angles, loose morals, kind, faithful, good-mother type, conscientious, and sincere. Ratings on the first two traits were transposed so that they would be comparable in direction to the others. The three pictures having the highest total score and the three having the lowest were selected. These are designated *HiMor* and *LoMor*, respectively.

The other two groups consisted of the three most ambiguous photographs (*HiAmb*) and two photographs (*LoAmb*) having low ambiguity. These groups differing in ambiguity were initially also selected so as to be neither high nor low on moral character. Only two pictures were included in the *LoAmb* group because other available photographs low in ambiguity were not neutral with respect to moral character. Ambiguity was measured by the *agreement index* (cf. 1), a Pearsonian *r* obtained by splitting the total group of judges into random halves and correlating the resulting pairs of mean ratings ($N = 35$ traits).

Rating procedure. Fifty-nine subjects rated each of the 11 photographs separately on a seven-point rating scale containing 20 personality characteristics expressed in everyday language (see Table 1). This scale was a shorter form of one used previously (2). After completing these ratings, the subjects rated their mothers on the same 20-item scale. On every trait, then, five scores were obtained for each judge: the mother-rating, *HiMor*, *LoMor*, *HiAmb*, *LoAmb*. The latter were obtained by averaging ratings of the several pictures in each of the four groups.

RESULTS

Extent of agreement among judges on mother-ratings. Hypothesis 1 asserts that judges markedly agree on ratings of their mothers. An agreement index computed for these ratings yielded an *r* of .96 ($N = 20$), strongly supporting the hypothesis. This *r* of .96 does not mean that agreement is so high that there is no individual variation among judges. It should be kept in mind that this frequently used index is based on the correspondence between the *average* ratings of one-half of the group of judges and the *average* ratings of the other half. The correlation is run over traits rather than persons; consequently, even when a high index is obtained, some variation among judges may be occurring. This is especially true when a large number of judges are included in the averages on which the index is based. For an extensive discussion of this index, see (1).

The "modal mother." To obtain a picture of what the stereotype of mother looked like, the mean ratings for all judges on each trait were computed. This profile of ratings is termed the *modal mother*. Averages alone, however, do not give a complete representation. Variability in ratings is also important; on some traits the judges are highly consistent in giving a similar rating to their mothers, e.g., *kind*, whereas for others, the judges range over the entire rating scale, e.g., *moody*. Figure 1 gives a composite picture of the mother-concepts of all judges in terms of means and variability, the latter being represented by quartiles. Traits highly characteristic of the modal mother are: kind, sincere, conscientious, refined, feminine, good mixer. Two traits conspicuously lacking are: conceited, loose morals.

Degree of similarity of the photograph groups to the modal mother. As a check upon whether the *HiMor* and *LoMor* photographs actually differed in the ease with which mother-traits could be assigned, the mean ratings for each photograph group were correlated with the mean ratings for the modal mother, with $N = 20$ traits. Modal mother ratings were found to correlate with *HiMor* +.88, and with *LoMor*, −.68. Thus, it may safely be assumed that these two photograph groups differ inversely in their similarity to the judges' modal mother.

With respect to the *HiAmb* and *LoAmb* groups, the original intent was to obtain pictures which were uncorrelated with mother-ratings. For an initial sample of 20 judges, modal mother ratings correlated .12 with *HiAmb* and −.03 with *LoAmb*, and on this basis the photographs selected were believed to be satisfactory. However, for the 59 judges finally employed, modal mother ratings correlated +.42 with *HiAmb* and +.60 with *LoAmb*, indicating that these photograph groups tended to yield impressions more

TABLE 1

ATTRIBUTION OF MOTHER-CONCEPTS TO PHOTOGRAPHS OF WOMEN

(Attribution is measured by computing a Pearsonian r between mother-ratings and photograph ratings)

Trait rated	Photograph Groups			
	HiMor	*LoMor*	*HiAmb*	*LoAmb*
Knows all the angles	.52**	−.07	.43**	.26*
Conceited	.47**	.21	.42**	.26*
Likeable	.42**	.12	.22*	.37**
Likes men's attention	.41**	.23*	.26*	.12
Easily hurt	.40**	.23*	.12	.15
Kind	.39**	−.05	.27*	.25*
Sharp-tongued	.38**	.05	.37**	.17
Loose morals	.31*	−.07	.18	.38**
Sincere	.30*	−.26*	.21	.33**
Conscientious	.30*	−.43**	.34**	.10
Feminine	.29*	.18	.15	.26*
Demanding	.29*	.10	.06	.22*
Moody	.22*	−.11	−.03	−.03
Good mixer	.21	.16	.26*	.41**
Sexually passionate	.19	.32**	.35**	.28*
Immature	.16	.24*	.30*	.29*
Gay	.12	.09	.29*	.23*
Refined	.08	−.10	.02	.00
Sexually attractive	−.05	−.03	.10	.00
Talkative	−.07	.17	−.10	.12
Mean r	.27*	.05	.21	.21
Difference in r	.22		.00	

Note.—Using a one-tailed test with $N = 59$, $r = .32$ at the .01 level and $r = .22$ at the .05 level.

* $p < .05$.

** $p < .01$.

like the modal mother than would be true of photographs having no positive cues for moral character. The important consideration, for purposes of control, however, is that the two groups have approximately similar (but not necessarily zero) relationships to the modal mother, and fortunately, the two groups do not differ greatly in this respect.

Attribution of mother-characteristics to photograph groups. The attribution occurring for the various photograph groups is shown in Table 1. A significant amount occurs for all photograph groups except *LoMor*. Thus, the hypothesis that attribution will occur for photographs of young women is, in general, confirmed. Also supported is the corollary that photographs having facial characteristics which more readily yield impressions of moral character will lead to a higher degree of attribution than photographs yielding impressions of low moral character. Thus, 15 of the 20 r's for the *HiMor* group are significant at the .05 level or better; whereas only 4 of the 20 r's for the *LoMor* group approach significance in the positive direction. The hypothesis that more ambiguous photographs will permit more attribution than unambiguous ones is, however, unfounded. There is no difference between the *HiAmb* and *LoAmb* group in the amount of attribution; both lead to a significant amount of attribution for about half the traits.

An examination of Table 1 fails to reveal a systematic explanation for the extent to which various *traits* permit attribution. Although a considerable range in r's is apparent, all attempts to find some rational interpretation of these differences have been unsuccessful. None of the following aspects of the personality characteristics covary systematically with attribution: content or meaning, objectivity or consistency of judgment, the degree to which they are characteristic of the modal mother, the variability of the mother-ratings. Perhaps the best interpretation of the variation in attribution for the different traits is in terms of sampling fluctuations.

One possible source of spurious correlation must be ruled out. One might argue that certain kinds of response sets or "number

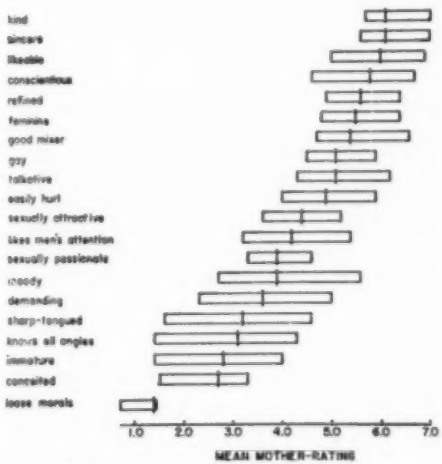

FIG. 1. THE MODAL-MOTHER PROFILE

The vertical line sectioning each bar represents the mean mother-rating for all judges. Ends of the bar represent the quartiles Q_1 and Q_3.

habits" may lead to spurious *r*'s, i.e., suppose that some judges have a response tendency to rate high for mother and for all photographs, and others have a similar tendency to rate low. Such response sets would tend to produce spurious correlation rather than correlation which may properly be designated attribution of mother-ratings. To check this possibility, all judges whose ratings of all photographs deviated on the average at least ± 0.5 of a scale point from the group mean of 4.0 were eliminated from the distribution, and *r*'s recomputed for the remaining 45 judges, on the first seven traits. The average change from the original sample to this sample of *r*'s was only +.01. Thus, response sets of this variety may safely be ruled out as a factor affecting the size of the obtained *r*'s.

DISCUSSION

Group agreement on mother-ratings, expressed by an agreement index of .96, reveals a surprising degree of stereotyping in the perception of one's mother. To the extent that mothers actually adopt the role pattern ascribed to them by our culture, this agreement is veridical. But certainly it is unlikely that mothers are really as similar as the ratings indicate. More probably, the conceptions of the judges are partly stereotyped to conform to cultural expectations. In spite of the marked agreement among judges, this variability, among them apparently has some significance as evidenced by the demonstration of a significant degree of attribution. Though variability in mother-ratings is small, it reflects "differences which make a difference."

As noted briefly in the introduction, at least two theoretical interpretations of the attribution of mother-concepts to young women are available. One of these is in terms of Sullivan's concept of parataxis (4, p. 101), a term applying to the process of reacting to others as if they represented a "significant other" from one's past. An adult of college age has in his lifetime associated more with his mother than with any other woman. During the long socialization period, conceptions of what the mother is like are developed. One may assume that these conceptions, whether veridical or autistic, will form a basis for generalized expectancies concerning the nature of other women. Since perceptual judgments, particularly complex social judgments of faces, are likely to be dependent upon expectancies, it is not surprising that a judge "sees" certain faces as having his mother's characteristics in some degree.

The other interpretation of attribution, not necessarily opposed to the interpretation in terms of parataxis, is that the judge, through all of his experience, not only with his mother, but with other people of both sexes and with life situations, has developed certain other-concepts concerning the nature of various classes of people. These might also be termed expectancies, but the difference between this interpretation and the preceding one is that, in our judgmental situation, these expectancies might affect the ratings of young women's photographs *and* of one's mother so as to produce a degree of similarity between them. In other words, the mother-concept might not have causal priority. Of course, mother-ratings do lean heavily in the direction of a cultural stereotype, which is not attributed to women in general, and also, attribution does not occur for photographs lacking the appropriate physiognomy (the *LoMor* group). These facts would appear to argue against this interpretation of attribution in terms of a general "concept of the other"; however, since the degree of attribution is not great, one could conceive of it as deriving from a general class construct in which both women and mother are included, but from which "bad" women (the *Lo Mor* group) are excluded. Present data do not permit us to choose between these two interpretations of our results. Whichever interpretation may prove correct, however, the process of attribution would appear to be a significant one for the understanding of interpersonal perception.

Incidentally, there appears to be no cogent reason why, if attribution occurs for pictures of women, it should not be expected to occur for women in life-situations. Not only is this likely for first impressions, but in view of the limited capacity of individuals accurately to judge the characteristics of people (5), attribution of mother-traits to women with whom one is acquainted may also occur.

SUMMARY AND CONCLUSIONS

On a seven-point scale consisting of 20 traits expressed in everyday language, 59 judges rated their mothers and four groups of photographs: *HiMor*, *LoMor*, *HiAmb*, and *LoAmb*.

The groups *HiMor* and *LoMor* consisted of photographs which a previous study had shown to represent extremes in the degree of *moral character* attributed to them, moral character being represented by a correlation cluster of such traits as kind, faithful, good-mother type, etc. The other two groups represented extremes in the degree to which judges were consistent with one another in the ratings they ascribed to the photographs, inconsistency being represented by *HiAmb*.

It was hypothesized that, on a particular trait, a judge's mean rating of women's faces would differ from the mean rating of the other judges in the same direction that his mother-concept deviates from the modal mother, e.g., if a judge rates his mother higher on the trait *sincere* than other judges rate their mothers, he will also rate young women higher on *sincere* than do other judges. This phenomenon, termed *attribution*, is measured for each trait by computing a Pearsonian r between mother-ratings and ratings of a given picture group. A large number of significant r's for 3 of the 4 photograph groups confirmed this general hypothesis. A corollary that the amount of attribution would be greatest for those photographs which are most easily perceived as having motherly traits was also confirmed. A second corollary, that attribution would vary directly with ambiguity of the photographs, was not confirmed; equal amounts occurred for photographs having many cues for judgment and those having few or contradictory cues.

Implications of the phenomenon of attribution for cognitive theory and for interpersonal relations were briefly discussed.

REFERENCES

1. SECORD, P. F., DUKES, W. F., & BEVAN, W. Personalities in faces: I. An experiment in social perceiving. *Genet. Psychol. Monogr.*, 1954, **49**, 231–279.
2. SECORD, P. F., & MUTHARD, J. E. Personalities in faces: II. Individual differences in the perception of women's faces. *J. abnorm. soc. Psychol.*, 1955, **50**, 238–242.
3. SECORD, P. F., & MUTHARD, J. E. Personalities in faces: IV. A descriptive analysis of the perception of women's faces and the identification of some physiognomic determinants. *J. Psychol.*, 1955, **39**, 269–278.
4. SULLIVAN, H. S. *Conceptions of modern psychiatry*. Washington: Wm. A. White Psychiatric Foundation, 1947.
5. TAFT, R. The ability to judge people. *Psychol. Bull.*, 1955, **52**, 1–23.

Received April 25, 1955.

A COMPARISON OF REPRODUCTION ERRORS MADE BY BRAIN-DAMAGED AND CONTROL PATIENTS ON A MEMORY-FOR-DESIGNS TEST[1,2]

H. J. WAHLER

State University of Iowa

THE task of reproducing abstract designs either from memory or by copying has a long history as a clinicodiagnostic procedure. It has been employed as an aid in identifying cases of organic brain damage both because of the long-standing observation that patients with cerebral disease tend to show significant impairment in retentive capacity and because of the belief, repeatedly expressed by clinicians, that qualitative features of performance in either the copying of designs or their reproduction from memory possess differential diagnostic significance.

Since the memory-for-designs subtests included in intelligence or deterioration test batteries are typically quite brief, several workers have devised longer series of designs with the intention of increasing reliability and the number of responses available for qualitative observation. Among the better known of these tests are the Benton Visual Retention Test (2), the Ellis Visual Designs (3) and the Graham-Kendall Memory-for-Designs Test (5).

Benton (2) and Graham and Kendall (5) have shown in independent studies that their respective tests were capable of correctly identifying about 50 per cent of the known brain-damaged *S*s in their samples with about 5 per cent of the control *S*s being misclassified. In both of these studies, total scores were employed as predictive measures without analyses of differences between the two groups in terms of specific types of errors.

With respect to qualitative features of reproductions, several authors have published their observations or compiled lists of the various features of graphic performance which they regarded as characteristic of brain-damaged *S*s (1, 7, 8, 10, 12). While such observations and lists have accumulated in the clinical literature over the past forty years, there has been virtually no research conducted to test their stability and generality.

Since there is clear evidence that brain-damaged *S*s tend to make more errors than controls on this type of task, the question is raised as to whether they make distinctively *characteristic* errors or whether they merely make *more* errors than controls. If it can be shown that brain-damaged *S*s tend to make characteristic types of errors, then a refined scoring system based on different types of errors may, with appropriate weights, increase the discriminatory power of a memory-for-designs test beyond that obtained with simple right-wrong scores.

The main purpose of the present investigation was to determine whether there are characteristic differences in the kinds of errors made by brain-damaged *S*s as compared with control *S*s in their respective performances on a memory-for-designs task.

PROCEDURE

Subjects

The brain-damaged *S*s used in this investigation were 48 patients in the neurological and neurosurgical services of the University Hospital and the Veterans Administration Hospital in Iowa City, Iowa. The 50 control *S*s were selected from these and other services of the two hospitals.

Criteria for Selecting Subjects

All *S*s were between the ages of 19 and 50 years, inclusive. An index of verbal IQ, consisting of the prorated sum of weighted scores on the Information and Comprehension subtests of the Wechsler-Bellevue Intelligence Scale, Form I, was obtained for each *S*. Only those *S*s were used in this study whose scores on these subtests were equivalent to a verbal IQ of 80 or higher. As shown in Table 1, the experimental and control groups were essentially equated on age and education.

The *S*s who were diagnosed as schizophrenic or manic-depressive or who, in the judgment of the examiner, were unable to cooperate were excluded. The

[1] Adapted from a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology, State University of Iowa. The author wishes to express his appreciation to Professors Arthur Benton, Paul Blommers, and Bertram Cohen for their valuable assistance.

[2] From the Veterans Administration, Iowa City, Iowa, where the author was employed when this study was undertaken.

TABLE 1

AGE, EDUCATION, WECHSLER SUBTEST AND VISUAL RETENTION TEST (VRT) SCORES FOR BRAIN-DAMAGED AND CONTROL *Ss*

	Group	Mean	Variance	M Diff.	*t*	*p*
Age	BD*	36.27	75.52	1.49	.88	.40
	C†	34.76	63.85			
Education	BD	10.04	7.37	.36	.68	.50
	C	10.40	6.00			
Information and Comprehension	BD	18.89	19.65	1.80	2.11	.05
	C	20.68	15.39			
VRT number correct	BD	5.33	4.22	1.63	4.32	.001
	C	6.96	2.64			
VRT total errors	BD	8.46	23.57	3.72	4.45	.001
	C	4.74	9.85			

* Brain-damaged; $N = 48$.
† Control; $N = 50$.

TABLE 2

DIAGNOSES OF BRAIN-DAMAGED AND CONTROL *Ss*

Brain-Damaged	*N*	Control	*N*
Cerebral trauma	11	Peptic ulcer	1
Neoplasm	11	Spinal disc pathology	10
Encephalitis	1	Polio	1
Cerebral thrombosis	1	Colitis	1
Cortical atrophy	6	Arthritis	3
Focal epilepsy	5	Neuritis	2
Organic brain disease	1	Pulmonary pathology	2
Vascular accident	5	Psychoneurosis	17
Frontal lobectomy	1	Sciatica	2
Subarachnoid hemorrhage	1	Tonsilitis	1
		Tension headaches	1
Pituitary adenoma	1	Maternity	2
Multiple sclerosis	2	Inner ear disease	1
Toxic state	1	Gastritis	1
Arteriosclerosis	1	Schizoid personality	1
		Anemia	1
		Sinusitis	1
		Undiagnosed (non-neurological)	2
Total	48		50

Ss with sensory or motor impairments severe enough to interfere with task performance were also excluded. The *Ss* for whom there was evidence of epilepsy, cortical pathology of any type, alcoholism, or endocrine pathology were not utilized as controls. The *Ss* who were assigned to the brain-damaged category had been diagnosed by a neurologist as having any one or more of the following conditions involving the cerebral cortex: neoplasms, vascular disorders, traumata, degenerative brain disease of unknown origin, abscesses or infections (e.g., paresis, encephalitis). Size or localization of lesion was not considered in the selection of cases. All cases in which the known onset of the disorder was earlier than 15 years of age were excluded. The diagnoses of the patients in both the brain-damaged and control groups are presented in Table 2.

Definitions of Error Categories and Scoring Procedure

A list of 28 different types of reproduction errors served as the basis for scoring. The 28 different types of errors were in turn organized into ten major categories as follows:

1. Complete omission of a major figure or failure to reproduce more than half of it.[3] (M)
2. Complete omission of a peripheral figure or failure to reproduce more than half of it.[3] (P)
3. Rotation of a major or peripheral figure (excluding the special case described in 4 below). (R)
4. Stabilization of a major figure; a special case of rotation in which any figure in an unstable position (i.e., resting on an angle) is reproduced as resting on a side. (StM)
5. Misplacement of major or peripheral figures in terms of their relative positions in the total design. (Mis)
6. Misplacement of a detail in a major figure. (MI)
7. Substitution of one figure for another, i.e., drawing a triangle instead of a square. Substitution is scored only when the error cannot be classified as perseveration. (Su)
8. Perseverative reproduction of a figure present in the immediately preceding design. Perseveration is also scored if the perseverated figure is again reproduced on the next design. (Per)
9. Distortion of size relationships—scored when the height of the peripheral figure (measured at point of maximal height) is greater than three-fifths the height of the larger of the two major figures (measured at point of maximal height) or when the height of the smaller of the two major figures is less than three-fifths the height of the larger. (Si)
10. Omission or addition of any details or parts of any figures and/or production of figures or part figures which were not present in the original design. (POA)

The total error score was defined as the sum of all errors regardless of type.

The objectivity of the error scores used in this study was evaluated by correlating (Pearson product-moment) the number of errors in each of the ten error categories assigned by one scorer to 29 randomly selected protocols with the number of errors assigned to corresponding categories by another independent scorer. These correlations ranged from .66 to .97. Category 7 had the lowest interscorer agreement ($r = .66$). Both categories 9 and 10 had interscorer correlations of .75. The remaining seven categories had interscorer correlations of .85 or higher.

Test Materials and Procedure

Benton's Visual Retention Test, Form C, and the Wechsler subtests were given to each *S* individually during the course of routine psychological examinations.

[3] Eight of the ten designs of the Visual Retention Test consist of two large figures, referred to as "major" figures, and one small figure referred to as a "peripheral" figure.

The standard procedure (Administration "A") was employed in the case of the Visual Retention Test. Benton's scoring criteria were used in determining number of designs correctly reproduced.

RESULTS

As shown in Table 1, the mean weighted Information and Comprehension subtest scores, which served as an index of verbal IQ, are significantly (.05 level) lower for the brain-damaged group than for the control group. This finding agrees with the results of a number of previous investigators (4, 6, 11), who showed that brain-damaged patients tend to perform more poorly on all tests, including verbal tests, than do controls.

The results with respect to Visual Retention Test performance are also presented in Table 1. It is evident that the mean number of correctly reproduced designs is significantly lower and the mean number of errors is significantly higher for the brain-damaged *Ss* than for the control *Ss*.

Analysis of Differences in Types of Reproduction Errors

The mean frequencies of the different types of errors made by the two groups and the ratios of the brain-damaged and control *Ss*' means are presented in Table 3. Inspection of Table 3 shows that the brain-damaged group made 1.78 times more total errors than the control group. With respect to most of the specific error categories, it will be noted that the ratios of the mean number of errors made by both groups are not appreciably different from the value of 1.78 for the ratio of total errors. Two exceptional categories in this regard, however, are size errors and rotational errors (excluding stabilization rotations) which show much higher ratios (6.00 and 4.00 respectively). These two types of errors occurred rather infrequently in the performances of the controls; when summed, they constitute only about 4 per cent of all the errors made by this group. They were much more frequently encountered in the performances of the brain-damaged patients, their sum constituting about 12 per cent of the errors made by the latter group.

TABLE 3

MEAN FREQUENCY OF DIFFERENT TYPES OF ERRORS IN BRAIN-DAMAGED AND CONTROL GROUPS

Type of Error	Brain-Damaged (N = 48) Mean	Control (N = 50) Mean	Ratio Between Means BD/C
Omission of major figures	.65	.44	1.48
Omission of peripheral figures	.62	.22	2.82
Rotations	.40	.10	4.00
Stabilization rotations	.77	.52	1.48
Misplacements	.96	.58	1.66
Misplacements of internal detail	.96	.58	1.66
Substitutions	1.21	.70	1.73
Perseverations	1.15	.84	1.37
Size errors	.60	.10	6.00
Partial omissions and additions	1.15	.66	1.74
Total	8.46	4.74	1.78

The question as to whether the two groups showed a significant over-all difference in the pattern of the several categories of errors was analyzed statistically by means of a trend analysis in the form of a Type I Design (9, p. 267). The F test for the significance of the group-by-error-categories interaction yielded an F of .26, which was too low to allow rejection of the hypothesis that the two trends were parallel. As was to be expected, the F test for the significance of the mean difference between the two groups in terms of vertical placement was significant (F = 8.48—.01 level with 1 and 30 degrees of freedom).[4]

A similar analysis was made of the performances of those *Ss* in each group who made "number correct" scores of 6 to 9 inclusive. This involved the exclusion of the 24 brain-damaged and the 5 control *Ss* who made scores of 5 or fewer correct reproductions and of the 2 control *Ss* who made perfect scores. By eliminating these subjects from both groups, the brain-damaged and control subgroups thus formed were roughly comparable

[4] Since the distributions based on individual measures were all J shaped, scores consisting of the sum of errors in each category for groups of three individuals were used in order to reduce the extreme skewness of the original data. These subgroup scores were used in order to approximate more closely the assumptions of normally distributed subject-by-error-category interaction effects and normally distributed subject means about group means for the between comparisons. Individuals were randomly assigned to these subgroups by means of a table of random numbers. In this analysis, the number of subgroups is 32; therefore, the scores of 3 × 32 or 96 individuals are represented. This device was employed in all analyses of variance mentioned in this study.

TABLE 4

MEAN FREQUENCY OF DIFFERENT TYPES OF ERRORS IN BRAIN-DAMAGED AND CONTROL *Ss* MAKING 6–9 CORRECT REPRODUCTIONS

Type of Error	Brain-Damaged (N = 24) Mean	Control (N = 43) Mean	Ratio Between Means BD/C
Omission of major figures	.25	.35	.71
Omission of peripheral figures	.21	.16	1.31
Rotations	.17	.05	3.40
Stabilization rotations	.62	.56	1.11
Misplacements	.46	.56	.82
Misplacements of internal detail	.62	.53	1.17
Substitutions	.83	.67	1.24
Perseverations	.79	.77	1.03
Size errors	.17	.09	1.89
Partial omissions and additions	.71	.51	1.39
Total	4.83	4.26	1.13

with respect to general level of performance. The mean "number correct" score of the brain-damaged patients scoring at the 6–9 level was 7.13 and that of the control patients scoring at the same level was 7.21.

The mean frequency of errors made in the different categories by the two groups with "number correct" scores of 6–9 and the ratios of their mean frequencies are presented in Table 4. The brain-damaged patients made 1.13 times as many total errors as the control patients. This finding is in accord with the slightly higher mean "number correct" score of the latter group. It will be noted that, while the ratios for most of the error categories do not differ appreciably from the ratio for total errors, size errors and rotational errors have the highest ratios (1.89 and 3.40) as was the case with the data presented in Table 3. It will be further noted, however, that these two types of errors occur less frequently than any of the other types. Moreover, a trend analysis of the performances of these subgroups scoring within the 6–9 range, in the form of a Type I Design yielded F values for the tests of significance of the groups-by-error-categories interaction and for vertical placement which were not significant.

DISCUSSION

The results indicate clearly that as far as level of performance on a memory-for-designs test is concerned, brain-damaged *Ss* tend to make more errors than control *Ss*. This finding is consistent with previous results and gives further support for the clinical usefulness of memory-for-designs tasks in the identification of cerebral pathology.

On the other hand, the failure to find significant differences in the patterns of types of errors made by the two groups does not support the clinical impressions which maintain that distinctive qualitative characteristics of performance appear in the graphic reproductions of brain-damaged *Ss* which differentiate them from other nosological groups. Although no significant differences in patterns of errors could be demonstrated, two specific types of error—size and rotational errors—appear from inspection to be sufficiently promising to warrant further study.

Several factors may be involved in this failure to secure results in agreement with the contentions of some clinical observers. First, it must be appreciated that "clinical impressions" constitute a rather unreliable guide for applied work although they can be a fruitful source of hypotheses. Frequently, such impressions derive from unsystematic observation in which the strikingly positive case is likely to be remembered and the numerous exceptions forgotten. Second, it is evident that in many instances these impressions have been formed on the basis of the performances of severely deteriorated patients. The findings of the present study suggest that when such patients are excluded from consideration and the sample is restricted to only moderately impaired *Ss*, discriminating differences in error characteristics fail to occur, leaving the difference in over-all level of performance as the sole distinguishing characteristic.

In addition, the failure to secure significant differences in patterns of performance may be related to the particular instrument, procedure, or criteria employed. It is possible that other types of designs or that one of the modifications in administration recently described by Benton (2) would be more successful than the standard procedure in eliciting differential response characteristics.

SUMMARY

The main purpose of this investigation was to determine whether there were characteristic

differences in the types of errors made by brain-damaged and control *Ss* in performance on a memory-for-designs task.

The brain-damaged and control samples used were comparable with respect to age and education. Only those *Ss* who gave evidence of intellectual functioning at the dull average or higher levels were retained. The performances of the brain-damaged and control groups on the Benton Visual Retention Test were compared in terms of total number of correct reproductions, total number of errors, and relative frequency of occurrence of ten different types of errors.

The mean total number of correct reproductions of the control *Ss* was found to be significantly higher and their mean number of errors significantly lower than the corresponding means of the brain-damaged *Ss*. No significant difference, however, between the two groups in the over-all patterns of the ten different types of errors could be demonstrated. On an inspectional basis, differences between the two groups with respect to specific types of errors suggested that size and rotational errors occurred with sufficiently greater frequency in the brain-damaged than the control group to warrant further study.

The finding of a significant difference in the number of correct reproductions made by the two groups confirms the results of previous investigators. The failure to find a significant difference in the patterns of the different types of errors made by the two groups does not support the contentions of a number of clinical observers. Possible reasons for the discrepancy between the present findings and the reports of these clinical workers were discussed.

REFERENCES

1. BENDER, LAURETTA. A visual motor gestalt test and its clinical use. *Amer. Orthopsychiat. Ass., Res. Monogr.*, 1938, No. 3.
2. BENTON, A. L. *The revised visual retention test: clinical and experimental applications.* New York: Psychological Corporation, 1955.
3. BRONNER, AUGUSTA F., HEALY, W., LOWE, GLADYS M., & SCHIMBERG, MYRA E. *A manual of individual mental tests and testing.* Boston: Little, Brown, 1927.
4. DELAY, J., PICHOT, P., DURSAP, R., & PERSE, J. L'aphasie latente. Son diagnostic dans l'arteriosclerose cerebrale au moyen d'un test de vocabulaire. *Rev. Neurologique*, 1950, **83**, 180–191.
5. GRAHAM, F. K., & KENDALL, B. S. Performance of brain-damaged cases on a memory-for-designs test. *J. abnorm. soc. Psychol.*, 1946, **41**, 303–314.
6. HEILBRUN, A. B., JR. Psychological test performance as a function of lateral localization of cerebral lesion. *J. comp. physiol. Psychol.*, in press.
7. HICKSON, W. J. The mental finger print. *Ann. Rep. Municipal Court of Chicago.* 1913–1915, **21**, 42–48.
8. HUTT, M. L. *A tentative guide for the administration of the Bender Gestalt Test.* Washington, Adjutant General School, U. S. Army, 1945.
9. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
10. SAPAS, E. Zeichnerishe Reproduktiones einfacher Figuren durch Geistkranke. *Schweiz. Arch. Neurol. Psychiat.*, 1918, **4**, 140–152.
11. SHAKOW, D., DOLKART, M. B., & GOLDMAN, R. The memory function in psychoses. *Dis. nerv. Syst.*, 1941, **2**, 43–48.
12. SHARP, AGNES A. The diagnostic significance of a visual memory drawing test. *J. abnorm. soc. Psychol.*, 1949, **44**, 517–527.

Received April 25, 1955.

ATTITUDE CHANGE THROUGH MODIFICATION OF ATTITUDE STRUCTURE[1, 2]

EARL R. CARLSON[3]
University of Michigan

THERE is wide agreement that "attitudes" are complex, in that they are composed of a number of components, characteristics, or dimensions. Psychologists generally agree, also, that changes in attitudes may come through the operation of different processes (4). There have been few studies, however, which have attempted to differentiate either components of attitudes or the processes of change experimentally. The present study is of this nature—an investigation of the importance of two independent sources of *affect* (the liking or favorableness-unfavorableness aspect) of an attitude.

The importance of the individual's system of general goals in life, or values, has been suggested by many theorists as of crucial significance in determining the affect associated with specific aspects of the person's experience. A recent study by Rosenberg (3), building upon and improving the methodology and conceptualization of earlier experimental studies (5, 6), demonstrated the relationship between attitudinal affect, i.e., the attitude position, or what is generally termed simply "attitude," and two components of "attitude structure." Rosenberg differentiated as structural components (*a*) the intensity of affect of a person's values, i.e., the expected satisfaction from these values, and (*b*) the perceived importance of the attitude object (situation, event) in leading to or blocking the attainment of the values. He measured "value satisfaction" and "perceived instrumentality" independently, and found each to correlate significantly with the attitude of the person. The best predictor of attitude, however, was found to be an index of "affective loading" which combined the two measures. The results clearly indicated that value satisfaction and perceived instrumental relations were separate and important components of the attitude structure.

Attitude change, it follows from this theory, should result from changes in either the expected satisfaction from goals, or in the instrumental relationship perceived between the attitude object and the goal. The present study was designed to test this latter hypothesis—that changes in attitude (affect) result from altering perceptions of the attitude object (situation or event) as leading to the attainment of valued goals. This hypothesis is consistent with experimental findings on animal and human learning, but has not been clearly tested in the area of social attitudes.

A second important source of affect is hypothesized to be mediated through generalization from affect associated with related attitude objects or situations. Following principles of generalization developed in learning studies, tests were made of two further hypotheses: (*a*) attitude change generalizes to related issues, and (*b*) the degree of generalization is a function of the similarity of the attitude objects or situations.

METHOD

The experimental procedure of the study was designed to increase the subjects' perception that "allowing Negroes to move into White neighborhoods" would be a means for attaining four specified goals, or values. The attitude issue of Negro housing segregation was selected as the vehicle for studying attitude and attitude structure change, since the subjects varied widely in attitude on the issue, the attitude was relatively stable, yet some change could be produced using a limited experimental procedure. Attitudes on five related issues were also measured in order to test for generalization of attitude change.

The four values discussed in the experimental communication were (*a*) American prestige in other countries, (*b*) protection of property values, (*c*) equal opportunity for personal development, and (*d*) being experienced, broad-minded, and worldly-wise. These values were selected on the basis of their probable importance for this issue.

The data were collected in three major stages. Early

[1] This study is one of a series undertaken by the Attitude Change Project at the University of Michigan. It was supported in part by the U. S. Air Force under Contract AF 33(038)-26646, monitored by the Human Resources Research Institute.

[2] The present paper is based upon a thesis submitted to the Department of Psychology of the University of Michigan in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer wishes to express his appreciation to Dr. Helen Peak for her generous advice and assistance given throughout the study.

[3] Now at Michigan State College.

in the semester experimental and control *Ss* were given attitude and value measures during regular class periods. Code numbers were used to maintain anonymity for the *Ss*. The experimental change procedure was administered to the experimental *Ss* in two parts: the first consisted of an assignment given by the instructor to be completed outside of class, one to two weeks after the initial measures were obtained; the second, five to ten days later, consisted of a prepared discussion given orally by the experimenter. After a further interval of three weeks the attitude and value measures were readministered to both experimental and control *Ss*. Finally, a follow-up questionnaire was given a few days later to obtain judgments of the perceived similarities between Negroes, Jews, and Mexicans.

In order to minimize the *Ss*' awareness of the nature of the study, the change procedure was presented as a separate experiment by a person not involved in the measurement of attitudes and value. The follow-up questionnaire indicated that *Ss* did not attempt to distort their responses deliberately.

Attitude measures. The measures of attitude toward the primary issue of study, Negro housing segregation, and the five related attitude issues required the *S* to rank six alternatives (ranging from "I am completely opposed to allowing Negroes into White neighborhoods" to "I am completely in favor of allowing Negroes into White neighborhoods") in terms of the degree to which they represented the person's own opinion. The measure of the person's attitude position used in the analysis was coded directly from the *S*'s first choice of the six alternatives. The measure of attitude change was obtained through coding the first three ranks given of the six alternative positions; using only transitive rankings, the three ranks gave a 16-point scale, ranging from position 1 (subject choosing alternatives in the order ABC), to position 2 (BAC), position 3 (BCA), and on to position 16 (FED). The Negro housing segregation measure had an uncorrected reliability of .67 for the control group.

To test the hypotheses concerning generalization of attitude change, five attitude issues similar to the Negro housing segregation issue on two dimensions were also measured in the same manner before and after the change procedure, these being (*a*) Mexican housing segregation, (*b*) Jewish housing segregation, (*c*) command of White enlisted men by Negro officers, (*d*) command of non-Mexican enlisted men by Mexican officers, and (*e*) command of non-Jewish enlisted men by Jewish officers.

Value measures. The measures of "value satisfaction" and "perceived instrumentality" required the *Ss* to rate each of the 25 value items given in Table 1. Two items, using different words but identical in meaning, were used within the set to measure each of the four areas selected for experimental change. The four sets of experimental items were items 5 and 19, items 8 and 11, items 10 and 22, and items 6 and 17. The remaining 17 values were selected to cover the broad range of further general goals in life that individuals regard as important in some degree. The items were presented on cards in random order, with a randomly assigned number on the card being used to record judgments of the value.

The value satisfaction index was obtained from *Ss*' ratings of the extent to which each value represented goals from which the *S* gets, or would get, satisfaction. Each value was judged independently in terms of 11 categories on a graphic rating scale. The scale of positively-valued goals was defined by three descriptive phrases: "No satisfaction from the goal" (Category 0), "Medium satisfaction from the goal" (Category 5), and "Maximum satisfaction from the goal" (Category 10). One further category was for judgments of "Dissatisfaction from the goal." The index of value satisfaction for the four experimental values was then obtained through arithmetic summation of the ratings of the eight items, coding "dissatisfaction" ratings minus 5. This index yielded an uncorrected test-etesrt reliability of .79.

The measure of "perceived instrumentality" required the *Ss* to judge each value in terms of the "probability that the goals will be attained by allowing

TABLE 1

VALUE ITEMS USED IN THE SATISFACTION AND INSTRUMENTALITY RATINGS

Value Item
1. Everyone being assured of a good standard of living
2. People sticking to their own groups
3. People looking out for the welfare of others
4. Being looked up to by others
5. America having high prestige in other countries
6. Being well-rounded, enlightened and sophisticated about life
7. Serving the interests of the group to which one belongs
8. Security of the value of one's real estate
9. Having power and authority over people
10. All persons having the chance to realize their potentialities
11. Having the value of property well-protected
12. Self-discipline—overcoming my irrational emotions and desires
13. All human beings having equal rights
14. Complying with the wishes of persons in authority
15. Being like others in general; having the same interests, opinions and ways of behaving as other people have
16. People having strict moral standards
17. Being a person who is experienced, broadminded, and worldly-wise
18. The open expression of disagreement between people
19. People in other nations respecting our principles and standards
20. Being allowed to maintain the privacy of one's opinions and beliefs
21. Being with other people; socializing
22. Everyone having opportunity to develop himself and his capacities
23. Letting others make their own decisions
24. People being strongly patriotic
25. Not being ashamed of one's own feelings and behavior

Negroes to move into White neighborhoods." The judgments were made on an 11-point graphic rating scale defined by three descriptive phrases: "Maximum probability that the goal will be blocked by allowing Negroes to move into White neighborhoods" (Category −5), "Allowing Negroes to move into White neighborhoods is irrelevant to achieving the goal" (Category 0), and "Maximum probability that the goal will be achieved by allowing Negroes to move into White neighborhoods" (Category +5). The index of perceived instrumentality was a summation of the ratings of the eight experimental value items, and had an uncorrected test-retest reliability of .66.

An index of "affective loading," used by Rosenberg to predict attitude, was computed separately for the total set of 25 value items, the eight experimental items, and the 17 nonexperimental value items. Since, according to the theory, attitude is a function of both value satisfaction and perceived instrumentality, attitude change should covary with change in an index combining measures of these separate components. The index of affective loading was computed by algebraically summing, over the set of values, the products of the satisfaction and instrumentality ratings of each value.

Change procedure. The change procedure was designed to increase the *Ss*' awareness that "allowing Negroes to move into White neighborhoods" would tend to bring about the attainment of the four values. This procedure consisted of two parts: (*a*) a task which was presented as a "Test of Objectivity," and (*b*) a verbal discussion which ostensibly reported the results of the test. The "Test of Objectivity," completed as an assignment by the students outside of class, was introduced as a test of one's "ability to take a certain viewpoint and support it logically and objectively regardless of your own viewpoint." It was presented as a small experiment illustrating the scientific method, studying whether men or women were better able to take an objective viewpoint in reasoning. The test required the *Ss* to support four propositions, each stating that "allowing Negroes to move into White neighborhoods" would lead to one of four values listed above.

In the second part of the change procedure the experimenter, in reporting the results of the study on objectivity in reasoning, presented a prepared discussion of the four propositions. The discussion, covering each of the values in 350 to 400 words, attempted to make clear several important and reasonable points which would be accepted by the *Ss*, and which demonstrated the instrumentality of "allowing Negroes to move into White neighborhoods" for attaining each of the four values.

Subjects. One hundred eighty-three students in eight introductory psychology sections at the University of Michigan participated as experimental *Ss*; 126 of these received both parts of the experimental procedure (the "Test of Objectivity" and the verbal discussion of the test results), and 58 received only one or the other of the two parts. The analyses reported in this paper are based upon the combined group of 183 experimental *Ss*.[4] Thirty-eight students in two classes participated as control subjects. The experimental and control groups were equated for initial attitude toward Negro housing segregation, and for both groups initial attitudes approximated a rectangular distribution.

RESULTS AND DISCUSSION

Effects of the Experimental Procedure

Before testing the basic hypotheses of the study, it must first be shown that the experimental procedure in fact altered both attitude and attitude structure as anticipated. The proportion of *Ss* changing attitude in a positive direction (toward less prejudice) on the Negro housing segregation issue was significantly greater for the experimental group than for the control group. Similarly, a significant proportion of experimental *Ss* changed in the predicted direction on instrumentality ratings of the eight experimental value items ($\chi^2 = 9.91$, $p < .01$).

Change in the value satisfaction ratings was not predicted, since the experimental procedure was directed only toward change in perceived instrumentalities, and no change was found. A significant proportion of experimental *Ss*, however, changed on the indices of affective loading based upon the eight experimental items ($\chi^2 = 17.79$, $p < .001$) and on the total set of 25 items ($\chi^2 = 4.19$, $p < .05$). Although not predicted, this change is consistent with the theory and with the rationale of the index, which is a direct function of instrumentality ratings as well as value satisfaction ratings. No systematic change was observed for the control group on the Negro segregation attitude scale or on any of the attitude structure measures.

It was anticipated prior to analyzing the data that *Ss* with different initial attitudes toward Negro housing segregation would react differently to the experimental procedure, but specific predictions were not made concerning the form of the relationships. No systematic differences were found for attitude structure changes, but the data clearly indicated a curvilinear relationship between initial attitude and attitude change for the experimental

[4] Separate analyses for the experimental subjects receiving both parts of the experimental procedure and for those receiving only one part demonstrated the same relationships throughout (1). The subjects receiving only one part generally showed less total change, as would be expected, but since the relationships between types of change were identical for both groups, the analyses reported here are based upon the combined group.

Ss, as shown in Figure 1. Tests comparing positive changes against the combination of no-change and negative-change *Ss* indicated that a significantly smaller proportion of *Ss* at position 1 (extreme prejudice) changed positively than *Ss* at positions 2 ($p < .05$) and 3 ($p < .05$); at position 4 the difference was at the .10 level. Similarly a significantly smaller proportion of *Ss* at position 5 changed positively than at positions 2 ($p < .01$), 3 ($p < .01$) and 4 ($p < .02$). The *Ss* at the moderate positions (positions 2, 3, and 4) were found to be significantly influenced by the experimental procedure, while extremely prejudiced (position 1) and extremely nonprejudiced (positions 5 and 6) persons were not found to change systematically. An analysis of change in the control group is not meaningful as a result of the small number of *Ss* at each initial attitude position.

A study of change on the five related attitude issues also supports this curvilinear relationship. Combining the data from all six scales showed that experimental *Ss* at position 1 were significantly less likely to change positively than *Ss* at position 2 ($p < .01$), position 3 ($p < .01$) and position 4 ($p < .01$), and that *Ss* at position 5 were also significantly less likely to change than those at position 2 ($p < .001$), position 3 ($p < .001$) and position 4 ($p < .05$). A further analysis of change for the generalization scales studied separately indicated basically the same curvilinear relationship for all six scales.

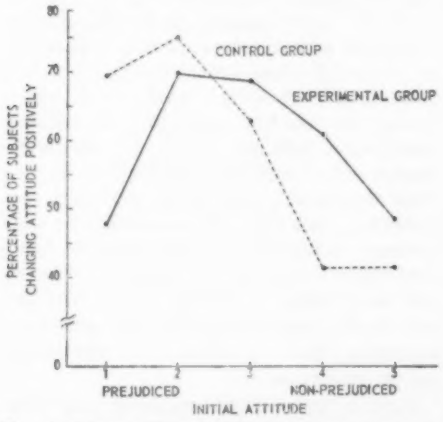

FIG. 1. RELATIONSHIP BETWEEN INITIAL ATTITUDE AND ATTITUDE CHANGE

For the control group, however, the more prejudiced *Ss* changed more in attitude than those at less prejudiced positions. A significantly larger proportion of *Ss* at positions 1 and 2 combined changed positively than *Ss* at positions 4 and 5 combined, suggesting the operation of other change influences during the period of the study.

What accounts for this curvilinear relationship between initial attitude and change? The smaller proportion of positive changes at the nonprejudiced positions can be explained by the fact that the *Ss* already largely agreed with the change communication, and by the ceiling placed on change by the measuring instrument itself. The finding for experimental *Ss* of fewer positive changes in attitude for prejudiced *Ss* than for those at intermediate positions can be accounted for in terms of the possible wide range on the attitude continuum over which *Ss* at this extreme position may be spread. Some prejudiced *Ss* may have actual attitudinal positions far below the cutting-point between positions 1 and 2, and the same amount of actual positive change would result in fewer position changes than for *Ss* initially at other positions.

These findings indicate the importance of considering the possible interaction of initial attitude on effects of attitude change techniques. In the present study the absence of attitude change for *Ss* at the extreme positions prevented tests of the major hypotheses for these *Ss*. Consequently the analysis of relationships between attitude structure change and attitude change, and the generalizations from the findings, must be restricted to persons with moderate initial attitudes toward the issue.

Relationships of Attitude Change to Attitude Structure Change

The hypothesis that *Ss* who changed in perceived instrumental relationships would also change in attitude was clearly supported by the data. Experimental *Ss* with moderate initial attitudes (positions 2, 3, and 4) who altered their perception of the relation of segregation to the values discussed also tended to change their attitude toward Negro housing segrega-

tion, this difference being significant at the .01 level ($\chi^2 = 24.01$, $n = 69$, $df = 9$).

Changes in the value satisfaction measures, if occurring, should covary with changes in attitude, but the amount of value satisfaction change was too slight to enable an adequate test of this relationship. Changes in attitude, however, were significantly related to the indices of affective loading, which reflected both instrumentality and value satisfaction changes. The relationship for the total set of 25 value items was significant at the .02 level ($\chi^2 = 5.27$, $n = 69$, $df = 1$), and the relationship for the eight experimental items was significant at the .05 level, measured in terms of a correlation coefficient ($r = .27$, $df = 68$), but not significant tested by chi-square analysis ($\chi^2 = 1.43$, $n = 70$, $p < .15$). A relationship found between attitude change and change in the index of affective loading for the 17 nonexperimental values ($\chi^2 = 5.47$, $n = 70$, $p < .01$) is also consistent with the theory, since the basic relationship between perceived instrumentality and attitude should hold for all values, whether experimentally manipulated or not.

What are the implications of these findings? This test does not account for all of the variance of attitude-affect change, certainly, but it has tested experimentally the significance of *one* important source of affect of an attitude. The role of perceived instrumental relations was demonstrated for *S*s with nonextreme attitudes, but was not tested for persons extremely prejudiced or extremely nonprejudiced. The underlying theory would predict that changes in value satisfaction would covary similarly with changes in attitude, but this hypothesis was not tested in the present study.[5]

Generalization of Attitude Change

The data demonstrated generalization of attitude change to certain related issues, but the hypothesis of a gradient of generalization varying as a function of the similarity of the attitude objects or situations was not confirmed. Comparison of the proportions of *S*s in the experimental group changing attitude positively, not changing, and changing negatively with changes in the control group indicated generalization of change to the Jewish housing segregation issue ($\chi^2 = 5.15$, $p < .03$), the Jewish officer issue ($\chi^2 = 8.43$, $p < .01$), and the Negro officer issue ($\chi^2 = 3.38$, $p < .07$). Because of the limited change in attitude for *S*s with extreme initial attitudes, the analysis here was also restricted to *S*s with initial positions 2, 3, and 4.

A sign-test analysis of positive changes compared to negative changes, i.e., discarding the no-change cases, rejected the null hypothesis (of no change in the positive direction) at the .05 level for all six attitude scales. A significant proportion of control *S*s summed over all six scales changed positively ($p < .05$), and a significant proportion changed positively on the two Mexican issues taken together; the other control group comparisons, however, were not statistically significant.

Judgments on a follow-up questionnaire item of "Which group do you believe is most like Negroes? —— Jews —— Mexicans" indicated that Mexicans were judged as "most like" Negroes by almost all *S*s, but, as reported, generalization of attitude change was not demonstrated for the two Mexican issues, though it was for the two Jewish issues. Actually significant proportions of experimental *S*s did change attitude positively on the Mexican scales, but the proportion changing in the control group, though less, was also significant. With the relatively small number of control *S*s the differences between groups were not statistically significant.

The failure to confirm the hypothesis of a gradient of generalization can be accounted for in terms of recognized limitations in the measure of "similarity" used in the study. Conceptual similarity between these issues is undoubtedly multidimensional, and the dimensions used by the *S*s in judging similarity may differ considerably from those over which generalization took place. Also several sources of evidence indicated that the Mexican issues were not well-structured or stable for these *S*s, and that changes occurring on them were due more to this lack of structure, and consequent instability, than to influence of the experimental procedure.

Further study of the process of generalization of change of social attitudes will depend upon systematic analyses of the gradients of

[5] This problem is currently under study in a program of research on attitude change processes by Dr. Helen Peak and her colleagues at the University of Michigan.

conceptual similarity involved. Methodology for handling this problem has not been thoroughly developed, though an approach formulated by Osgood (2), using a combination of associational and scaling procedures, may provide a systematic basis for such measurement and dimensionalizing of conceptual meaning.

SUMMARY

This experiment was designed to test the hypothesis that attitudes toward an object, or situation, may be changed through altering the person's perception of the significance of the object as a means for attaining valued goals. It sought further to test whether attitude change generalizes to related objects, and whether the degree of generalization is a function of the similarity of the objects.

One hundred eighty-three experimental *Ss* responded to three sets of measures before and after a change procedure: (*a*) measures of attitude toward "allowing Negroes to move into White neighborhoods" and five related issues, (*b*) a measure requiring ratings of 25 general values in terms of the expected satisfaction from each, and (*c*) a rating measure of the 25 values in terms of the probability that "allowing Negroes to move into White neighborhoods" would lead to, or block, the attainment of each of the values. Thirty-nine control *Ss* took the same measures but did not receive the change procedure. The change procedure was designed to increase the *Ss*' awareness that nonsegregation would lead to the attainment of each of four important values.

The experimental procedure changed perceptions of the role of Negro housing segregation in attaining the four values, and changed attitude toward Negro housing segregation for *Ss* with moderate initial attitude on the issue. Insufficient attitude change was produced in extremely prejudiced and extremely nonprejudiced subjects to permit tests of the hypotheses for these *Ss*. Satisfaction ratings were not affected by the change procedure.

Changes in attitude were related significantly to changes in perceived instrumental relationships, and to changes in an index based upon both satisfaction and instrumentality ratings. Attitude change generalized to three related attitude issues, but not to two further issues. The extent of generalization of attitude change was not shown to vary with the similarity of the attitude issues.

REFERENCES

1. CARLSON, E. R. Attitude change through modification of attitude structure. Unpublished doctor's dissertation, Univer. of Michigan, 1953.
2. OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197–237.
3. ROSENBERG, M. J. The experimental investigation of a value theory of attitude structure. Unpublished doctor's dissertation, Univer. of Michigan, 1953.
4. SARNOFF, I., & KATZ, D. The motivational bases of attitude change. *J. abnorm. soc. Psychol.*, 1954, **49**, 115–124.
5. SMITH, M. B. Personal values as determinants of a political attitude. *J. Psychol.*, 1949, **28**, 477–486.
6. WOODRUFF, A. D., & DIVESTA, F. J. The relationship between values, concepts, and attitudes. *Educ. psychol. Measmt*, 1948, **8**, 645–659.

Received May 5, 1955.

THE RELIABILITY OF PSYCHIATRIC DIAGNOSIS: A NEW LOOK

HERMANN O. SCHMIDT AND CHARLES P. FONDA[1]

Norwich State Hospital, Norwich, Connecticut

In a recent publication (14, p. 100), we asserted without documentation that "psychiatric diagnosis is too unreliable to permit derivation from our data of substantial conclusions. . . ." At the time that this statement was written, it was the authors' impression that such a proposition was widely accepted and rested securely upon a solid foundation of published evidence. Other writers have made use of this same rationalization for discrediting the use of psychiatrically diagnosed criterion groups (e.g., 7, 13), but during one of our *mea culpa* moods it occurred to us to ask, "Just how unreliable are psychiatric diagnoses anyhow?" The present communication is a report of our odyssey in search of an answer to that question.

Background

Summarizing such documentation as might have been found in the literature, we are chagrined to observe that precious little evidence has been published that could have justified our cavalier dismissal of psychiatric diagnosis as "too unreliable." It is true that some disagreements among psychiatrists have been reported, but these have not ordinarily involved the established nosology of their profession. The psychiatrists seem to have been about as successful as the rest of us in their attempts to use unequivocal language to communicate inferences about other people's behavior. That "marked disagreement" (5) and "little consensus" (6) have attended their efforts to appraise character traits and underlying motivations (12), or to recognize such vague entities as "primitive personality" or "nomadic personality" (1), is interesting but probably irrelevant to the problem of determining the reliability of the official diagnostic nomenclature.

Unfortunately, however, it is almost impossible to generalize about reliability from published reports of the studies which did use one of the conventional diagnostic systems. For example, the finding that about 40 per cent of psychiatric diagnoses required "major revision" (8) one year after discharge from a clinic may be due in an unknown degree to actual changes in patient status during the interval. Again, the showing that various psychiatrists use slightly different criteria for deciding between an organic and a psychogenic diagnosis, or between schizophrenia and manic-depressive psychosis (10), proves that these diagnoses must be something less than perfectly reliable, but it doesn't tell us how much less. Finally, the fact that medical students were able to achieve a "moderate" degree of success in arriving at independent agreement with the professor regarding the diagnoses of a half-dozen "classical" cases selected for class demonstration (15) leads only to a conclusion that some diagnoses can be reasonably reliable when applied to certain patients.

From our appraisal of studies in this area, however, it has been possible to generate some specifications for the design (2) of an ideal study of the reliability of psychiatric diagnoses. Five generalizations may be stated as follows: (*a*) The sample of patients should be representative of the population to whom the diagnoses will be applied. (*b*) The sample of diagnosticians should be representative of the population of psychiatrists who will do the diagnosing. (*c*) The diagnostic nomenclature should represent some nosological system in widespread use. (*d*) The diagnoses should be made independently and concurrently. (*e*) They should be obtained in a representative sample of clinics and hospitals.

Procedure

Each of 426 patients admitted to the Norwich State Hospital during a six-month period was diagnosed independently by two psychiatrists. The initial diagnosis was made by one of a group of eight psychiatric residents during the patient's first week in the hospital. The second diagnosis was made by one of a group of three chief psychiatrists during the patient's third week in the hospital. It might have been instructive to have had every patient diagnosed by each of the eight residents and by each of the three chief psychiatrists, but this cumbersome operation was not at-

[1] The authors wish to acknowledge their appreciation to the Superintendent, R. H. Kettle, M.D., whose interest and support made this study possible.

tempted. We shall refer hereinafter to the residents' diagnoses as *tentative*. They were communicated privately to one of the staff psychologists,[2] and care was taken to make sure that the tentative diagnoses were not communicated to the chief psychiatrists. The diagnoses made by the latter individuals are designated as *official*. They were made during a formal staff conference and recorded in the patient's chart.

At the time that the tentative diagnoses were made, the residents had the following data at hand: a short statement from the committing physician, the brief report of an admission interview by another physician, and observations during a physical examination and mental status interview. At the staff conference where the official diagnoses were made, the chief psychiatrist had access to reports covering all of the foregoing material, plus additional information from such sources as the social history, the psychological and neurological examinations, a public interview with the patient, and so on. Although the residents were also present at these staff conferences, the chief psychiatrists were never informed of the tentative diagnoses. During the period when data were being collected for the present study, the chiefs had agreed to take special precautions to avoid giving the residents any opportunity to contaminate the official verdicts.

Seven men and one woman constituted the group of residents who participated in this study. All but two were under 30 years of age, and six had received their medical degrees during the preceding three years. Only three of the eight degrees had been granted by medical schools in the United States. Six members of the group had had no previous psychiatric experience, whereas two had each held one-year psychiatric residencies in other hospitals. The tentative diagnoses were obtained from the residents during their first six months at this hospital, and the group as a whole can be characterized as young, foreign-trained, inexperienced, and psychiatrically naive.

All three of the men (the chief psychiatrists) who made the official diagnoses were certified by the American Board of Psychiatry and Neurology. One was the hospital Assistant Superintendent and two were Clinical Directors. All were mature, American-trained, widely experienced, and psychiatrically well-informed.

All diagnoses employed the nomenclature prescribed in the 1952 revision of the American Psychiatric Association's *Diagnostic and Statistical Manual* (4). This *Manual* lists 91 specific diagnoses available for use at this hospital, of which only 40 were actually used during the period covered by the present study. These classifications will hereinafter be designated as specific *subtype* diagnoses. Representative examples are "schizophrenic reaction, paranoid type" or "obsessive-compulsive reaction."

The subtype diagnoses were grouped for analysis into eleven classes of *disorder*, such as "schizophrenia," "psychoneurosis," "mental deficiency," etc. As a final step, these eleven disorders were further classified according to their membership in one of three major *categories*: organic, psychotic, or characterological.

[2] Grateful acknowledgment is due to Drs. Margaret B. Scales and Sidney A. Orgel for their gracious cooperation in gathering these data.

Contingency tables were prepared, showing for each resident and for the entire group the degree of concordance between the tentative and the official diagnoses with respect to the major categories. The proportions of official diagnoses for which the tentative diagnosis had correctly anticipated the specific subtype were then computed within each of the eleven disorders. Finally, every resident was assigned a set of "correct anticipation" scores. One set represented the proportions of each of the major categories correctly predicted; the other represented the proportions of subtype diagnoses correctly anticipated within each major category. Each set of scores was submitted separately to an analysis of variance.

RESULTS

Table 1 shows that the tentative and the official diagnoses were in agreement as to the major category in 84 per cent of the 426 cases. A contingency coefficient of .714 computed from these data indicates that a high level of reliable discrimination (3) had been achieved among the three diagnostic categories. (The maximum possible coefficient from a 3 × 3 table is .817; the obtained value is roughly equivalent to a Pearson r of .90 and is highly significant. Chi square is 443.55, and p is well beyond the .001 level.)

What Meehl and Rosen (9) have called "base rates" can be seen along the marginal per cent totals of Table 1. Thus, the organic category is used in about 45 per cent, the psychotic in 38 per cent, and the characterological in 17 per cent of the official diagnoses. That the residents may have acquired a reasonably reliable impressionistic knowledge of these proportions is shown by the fact that they employed the organic category in 47 per cent of cases, psychotic in 35 per cent, and characterological in 18 per cent. The value of the above chi square, however, indicates that mere knowledge of these institutional base

TABLE 1

OVER-ALL AGREEMENT BETWEEN PSYCHIATRIC RESIDENTS AND CHIEF PSYCHIATRISTS WITH RESPECT TO MAJOR CATEGORY OF DIAGNOSIS

		Chief Psychiatrists			Total	
		Organic	Psychotic	Character	*N*	Per cent
Residents	Character	3	21	51	75	18
	Psychotic	12	128	9	149	35
	Organic	178	12	12	202	47
Total	*N*	193	161	72	426	
	Per cent	45	38	17		100

TABLE 2

TETRACHORIC CORRELATIONS BETWEEN TENTATIVE AND OFFICIAL DIAGNOSES BY MAJOR CATEGORY

Resident	N	r_t Organic	Psychotic	Characterological
A	31	>.95	>.95	.94
B	58	.93	.95	.95
C	56	.95	.93	.92
D	63	>.95	.88	.90
E	67	.94	.87	.80
F	30	.92	>.95	.80
G	59	.95	.87	.72
H	62	.87	.82	.60
Group	426	>.95	.92	.85

rates alone could not have enabled the residents to produce the results obtained.

Table 2 shows the relative reliabilities of discrimination achieved for the three major categories by each of the residents. The strength of relationship between tentative and official diagnoses is indicated by tetrachoric coefficients of correlation. These coefficients were computed from fourfold tables giving the numbers of concordant and discordant answers by residents and chief psychiatrists to the question, "Does this patient have an *X* diagnosis?" When the question was, "Does this patient have an *organic* diagnosis?," the r_t for the group is above .95, and no individual's r_t is below .87. For the *psychotic* category, the group r_t is .92, and the lowest individual coefficient is .82. The group r_t for the *characterological* is .85 and one individual's coefficient is as low as .60. While all of these coefficients represent statistically significant relationships, and while the differences between the major categories are highly reliable, the differences *between residents* could easily be due to chance. This point will be developed later.

The reliability of the diagnosis of schizophrenia is of particular interest to many investigators (16), especially to those who wish to use diagnosed schizophrenics in criterion groups (7). If we were to consider that pairs of observers had been asked to answer the question, "Is this patient schizophrenic?" regarding all 426 cases, their answers could be tabulated as in Table 3. This table shows that residents and staff were in agreement regarding

TABLE 3

AGREEMENT BETWEEN RESIDENTS AND CHIEF PSYCHIATRISTS IN THE IDENTIFICATION OF SCHIZOPHRENICS AMONG ALL PATIENTS

	Chief Psychiatrists		Total
	Schiz.	Nonschiz.	
Residents			
Nonschiz.	19	287	306
Schiz.	99	21	120
Total	118	308	426

386 or 91 per cent of the total number of patients. The tetrachoric r computed from this table is .95, indicating a very substantial degree of reliability. It is widely known, however, that the magnitude of a reliability coefficient is a function of the range of variability represented. If we include in the fourfold table only those patients who had been considered psychotic by a resident, a chief psychiatrist, or both, and tabulate answers to the question, "Is this psychotic patient schizophrenic?" we obtain the tabulation shown in Table 4. Here the percentage of cases in which the resident and staff psychiatrist agree drops to 78 per cent, and the corresponding r_t becomes .73. The lower coefficient reflects the restriction in range produced by the elimination of clear-cut nonpsychotics; the effect of such elimination is to increase the difficulty of discriminating between schizophrenics and nonschizophrenics. Nevertheless, the lower value is still significantly better than chance and gives evidence of a minimally adequate degree of reliability. (Chi square is 48.46; p is less than .001.)

When the data are analyzed in terms of the

TABLE 4

AGREEMENT BETWEEN RESIDENTS AND CHIEF PSYCHIATRISTS IN THE IDENTIFICATION OF SCHIZOPHRENICS AMONG PSYCHOTIC PATIENTS

	Chief Psychiatrists		Total
	Schiz.	Nonschiz.	
Residents			
Nonschiz.	19	43	62
Schiz.	99	21	120
Total	118	64	182

TABLE 5

PROPORTIONS OF OFFICIAL DIAGNOSES CORRECTLY ANTICIPATED IN THE RESIDENTS' TENTATIVE DIAGNOSES

Classification	Official Diagnoses	Residents' Diagnoses Agreement as to:			
		Major category		Specific subtype	
	N	N	p	N	p
Organic	193	178	.92	142	.74
Acute brain syndrome	66			45	.68
Chronic brain syndrome	115			92	.80
Mental deficiency	12			5	.42
Psychotic	161	128	.80	75	.47
Involutional	14			8	.57
Affective	20			7	.35
Schizophrenic	118			60	.51
Unclassified	9			—	—
Characterological	72	51	.71	17	.24
Neurosis	25			4	.16
Personality pattern	12			1	.08
Personality trait	16			1	.06
Sociopathic	19			11	.58
Total	426	357	.84	234	.55

percentage of official diagnoses correctly anticipated in the residents' tentative diagnoses, as in Table 5, we find 84 per cent correct for the major categories, as against 55 per cent for the specific subtypes. The difference between these two correlated proportions is 6.9 times its standard error and is highly significant. The major category had been correctly anticipated in 92 per cent of the organic, 80 per cent of the psychotic, and 71 per cent of the characterological cases. There were 193 official diagnoses in the organic category. Of these, 74 per cent had been correctly predicted as to subtype. The prediction of subtype diagnosis was more efficient for some organic disorders than for others: in the chronic brain syndrome, 80 per cent of the subtypes had been correctly anticipated, whereas in only 47 per cent of the officially diagnosed cases of mental deficiency were the subtype predictions correct.[3] Of the 128 cases in the psychotic category, the subtypes had been correctly predicted in 47 per cent. When the official diagnosis was schizophrenia, the tentative subtype diagnoses had been correct in 51 per cent. The residents made their poorest showing for the 51 patients in the characterological group, where only 24 per cent of the subtype diagnoses had been accurately predicted.

[3] Mental deficiency has been included with the organic cases for the sake of simplicity although the *Manual* lists this disorder as a separate category.

TABLE 6

RESIDENTS' ACHIEVEMENT SCORES FOR CORRECT ANTICIPATION OF OFFICIAL DIAGNOSES

Resident	N	Diagnosis by					
		Major category*			Specific subtype†		
		Org	Psych	Char	Org	Psych	Char
A	31	92	92	80	85	23	20
B	57	88	85	80	69	50	12
C	56	88	85	83	67	58	0
D	63	96	84	77	76	56	39
E	67	94	73	62	70	38	0
F	30	100	67	70	79	33	50
G	59	94	69	57	78	50	14
H	62	88	73	43	72	43	43

* Between individuals $F = .76$ $p > .05$; between categories $F = 14.94$ $p < .001$.

† Between individuals $F = 1.98$ $p > .05$; between subtypes $F = 27.21$ $p < .001$.

Carrying the analysis of proportions of correctly anticipated diagnoses one step further, Table 6 gives the correct-anticipation achievement scores of the eight residents for both major category and subtype predictions. Two analyses of variance were performed for between-category differences in the data of Table 6, one for category predictions and the other for subtype predictions. Results of these analyses show that it would be impossible to predict any resident's score in one category from knowledge of his achievement in the other two. Variability of the scores is such as to preclude a demonstration of consistent differences among the residents with respect to diagnostic acuity. On the other hand, the data do reflect substantial differences in level of achievement between the various categories and thus indicate that official diagnoses are significantly more difficult to predict in some categories than in others.

DISCUSSION AND CONCLUSIONS

The study reported here is not an ideal one. Although it fully satisfies the requirements that the official nomenclature be employed and that the diagnoses be made independently and concurrently, it falls short of being adequately representative of psychiatrists-in-general to the extent that the sample is biased

by the presence of inexperienced residents. It can be pointed out, however, that it was impossible to demonstrate consistent individual differences between residents in the ability to anticipate official diagnoses, despite the fact that some had had considerably more training and experience than others. Moreover, the "nationally known" psychiatrists in Ash's (1) study had been unable to reach better than 45 per cent agreement as to subtypes (mainly characterological), whereas our residents anticipated 55 per cent of the official subtype diagnoses (mainly organic and psychotic) correctly. The point is that bias undoubtedly exists in our sample of psychiatrists and probably has had the effect of reducing the reliability coefficients somewhat. Our sample of patients is large enough to be fairly representative of state hospital populations, but no attempt was made to evaluate interinstitution diagnostic reliability. Such an evaluation is urgently needed but would have demanded more facilities than our modest resources provide.

When reliability is expressed in terms of a Pearson or a tetrachoric r, the coefficient must approximate .90 in order to insure 80 per cent agreement between two independent measures in a fourfold table. This statement rests upon the assumption that the two variables are continuously and normally distributed and the regression is linear. It is not certain that our data satisfy these assumptions in every detail, but it is clear from our results that a very substantial degree of agreement has been achieved for the two diagnoses which account for more than half of the admissions to this hospital. Thus, when an official diagnosis is either chronic brain syndrome or schizophrenia, the odds are four to one that an independent tentative diagnosis would be in agreement. On the other hand, the situation with respect to the subtype diagnoses in most of the disorders is far from satisfactory, although in another study conducted at this hospital, Orgel (11) found better than 90 per cent agreement among three psychologists who were asked to separate hebephrenic and paranoid schizophrenics from all other patients on the basis of blind diagnoses made solely from the patients' charts. The psychologists used an experimental set of behavioral criteria somewhat more detailed than those contained in the *Diagnostic Manual* (4). Less than 50 per cent of the cases selected in this way had been similarly classified by the official diagnoses. It also appears that the reliability of psychiatric diagnoses diminishes as the frequency of their incidence decreases. The most unsatisfactory diagnoses from the point of view of reliability are those in the characterological group. Since these are also the furthest removed from possible physiological or medical complications, it is likely that improvement in both the reliability and the usefulness of character diagnoses must await the development of more valid taxonomic methods. It seems reasonable to suppose that psychologists will be able to make substantial contributions to improvement in this area.

SUMMARY

Each of 426 state hospital patients was diagnosed independently by pairs of psychiatrists using the official psychiatric nomenclature. Reliability of the schizophrenic diagnosis was indicated by values of r_t between .73 and .95, depending upon the range of discrimination required. Classification of the diagnoses into three major categories—organic, psychotic, characterological—showed that about four-fifths of such classifications by one psychiatrist were confirmed by another. Agreement with respect to diagnoses of the specific subtype of a disorder occurred in only about half of the cases and was almost absent in cases involving personality pattern-and-trait disorders and the psychoneuroses. It is concluded that satisfactory reliability has been demonstrated for some of the psychiatric diagnoses, but that this carries no implications regarding their semantic validity or usefulness.

REFERENCES

1. ASH, P. The reliability of psychiatric diagnosis. *J. abnorm. soc. Psychol.*, 1949, **44**, 272–276.
2. BRUNSWIK, E. *Systematic and representative design of psychological experiments with results in physical and social perception.* Berkeley, Calif.: Univer. of California Press, 1947.
3. DAVENPORT, BEVERLY F. The semantic validity of TAT interpretations. *J. consult. Psychol.*, 1952, **16**, 171–175.
4. *Diagnostic and statistical manual. Mental disorders.* Washington, D. C.: American Psychiatric Association, 1952.
5. DOERING, C. R., & RAYMOND, A. F. Reliability of observation in psychiatric and related charac-

teristics. *Amer. J. Orthopsychiat.*, 1934, **4**, 249–257.

6. ELKIN, F. Specialists interpret the case of Harry Holzer. *J. abnorm. soc. Psychol.*, 1947, **42**, 99–111.
7. KING, G. F. Research with neuropsychiatric samples. *J. Psychol.*, 1954, **38**, 383–387.
8. MASSERMAN, J. H., & CARMICHAEL, H. T. Diagnosis and prognosis in psychiatry: with a follow-up study of the results of short-term general hospital therapy of psychiatric cases. *J. ment. Sci.*, 1939, **84**, 893–946.
9. MEEHL, P. E., & ROSEN, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.*, 1955, **52**, 194–216.
10. MEHLMAN, B. The reliability of psychiatric diagnosis. *J. abnorm. soc. Psychol.*, 1952, **47**, 577–578.
11. ORGEL, S. A. Clustering of verbal associates in schizophrenia and chronic brain syndrome. Unpublished doctor's dissertation, Univer. of Connecticut, 1956.
12. RAINES, G. N., & ROHRER, J. H. The operational matrix of psychiatric practice. I. Consistency and variability in interview impressions of different psychiatrists. *Amer. J. Psychiat.*, 1955, **110**, 721–733.
13. RUBIN, H. The MMPI as a diagnostic aid in a veterans hospital. *J. consult. Psychol.*, 1948, **12**, 251–254.
14. SCHMIDT, H. O., FONDA, C. P., & LESTER, J. R. Rorschach behavior as an index of color anaesthesia. *J. Psychol.*, 1955, **40**, 95–102.
15. SEEMAN, W. Psychiatric diagnosis. An investigation of interperson-reliability after didactic instruction. *J. nerv. ment. Dis.*, 1953, **118**, 541–544.
16. WOOLLEY, L. F. Experiential factors essential to the development of schizophrenia. In P. H. Hoch, & J. Zubin (Eds.), *Current problems in psychiatric diagnosis*. New York: Grune & Stratton, 1953. Ch. 13.

Received September 26, 1955. Prior publication.

CASE REPORTS

MOTOR MANIFESTATION OF CONFLICT IN INTERVIEW: A CASE STUDY[1]

ROBERT B. MALMO, A. ARTHUR SMITH, AND WERNER A. KOHLMEYER[2]
Allan Memorial Institute of Psychiatry and McGill University

MUSCLE-POTENTIAL recording from patients undergoing interview has proved useful in the study of symptom mechanisms (1, 3). Recently, Shagass and Malmo (6), using the method of continuous electromyographic (EMG) recording during psychiatric interview, were able to show a clear relationship between interview content and localized patterns of muscle tension. In brief, they found high forearm tension associated with conflicts over hostility, and high leg tension associated with sexual conflicts.

Content analysis in their study was done independently by two persons using a rating technique. Fair agreement between raters was found; but, as a check on the possibility of rater bias, it seemed worthwhile to repeat the experiment with two new raters and with a different therapist. Verification of the findings under changed conditions would do much to confirm the validity of the phenomena.

Case History of Subject

Mrs. H., a 33-year old widow, was treated as inpatient in the Allan Memorial Institute for five weeks. On admission her symptoms consisted in feelings of "tension" and anxiety, fainting spells without complete loss of consciousness, whistling in her left ear, and sensations of squeezing and tightness in her pelvis.

A precipitating factor in her illness was the death of her husband seven months prior to her admission. After her husband's death, she left her home city to work in Montreal, leaving behind her baby daughter. In Montreal she had a love affair which resulted in an apparent pregnancy which she tried to terminate with the help of an abortionist. Subsequently, she suffered a hemorrhage and was treated in the Royal Victoria Hospial. Her pregnancy tests were ambiguous and a curettage was carried out. The patient was seen by a psychiatrist and referred for treatment in the Institute.

Family history. The patient's father, 62, a pipe fitter by trade, was an alcoholic who frightened his family often with his drinking, although he was not actually violent when intoxicated. In the frequent quarrels between the parents, the patient's sympathy lay with her mother. The mother died of hypertension at age 52 when the patient was 25. The father had a bad stroke recently and is threatened with cirrhosis. The patient has no close contact with her father at present.

The patient has one sister, six years her senior, whom she describes as a rigid, unfriendly person. The sister was an object of strong hostility for the patient, although much of this was unconscious, and there was considerable ambivalence, which is clear in the following verbatim excerpt from the interview: "Sometimes I used to feel like screaming at my sister—I can remember that—and yet, I'd do anything for her, and she'd do anything for me. I don't know why my sister makes me so mad. I don't want to feel that way." After her husband's death, the patient left her child with her sister, against whom she now expresses resentment for taking away the child's affection.

Personal history. The patient was born in England. Coming to Canada at the age of two, she began school at the age of six, continuing to grade seven which she reached at the age of twelve. Soon afterward she began work in a department store as parcel girl. Later she taught herself to type and has worked in a number of jobs as stenographer.

Her menarche was not particularly stressful but the patient developed a tomboyish attitude early in adolescence, and she appeared

[1] This investigation was supported by the Medical Research and Development Division, Office of the Surgeon General, Department of the U. S. Army, under Contract No. DA-49-007-MD-70.

[2] The authors wish to acknowledge the invaluable assistance of Dr. Michael Kaye who carried out the interview, of Mr. R. E. Quilter who assisted in taking the records, and of Mrs. David Hubel who measured the records.

to find it difficult to adjust to the feminine role. This situation changed gradually, however, and the patient began to enjoy the company of men, especially men considerably older than herself.

At 24, she married a man of 29 whom, after three months of marriage, she discovered to be a bigamist. The marriage was legally annulled. Two years later the patient married a man of 31, whom she said falsely represented himself as a major in the Army and millionaire's son. The patient learned some time later that the first part of this marriage had also been bigamous, because the husband's divorce of his first wife was not completed until after he married the patient. Fear of pregnancy seemed an important factor in making this second marriage unsatisfactory. The husband, moreover, was prone to outbursts of rage and violence, and he was an irresponsible person who did not provide any security for the patient. It was in the setting of these unfavorable marital conditions that the patient had a child after five years of marriage. Because of her husband's death she is presently unable to provide a home for her child.

Diagnostic data. At the time she was seen, the patient was having an affair with a married man who, though promising to divorce his wife, had shown very little evidence of serious intention to do so. This and other promiscuous relationships with men indicated that the patient was suffering from sexual conflicts. It is perhaps significant to note that she recognized her tension and feelings in her pelvis as sexual in origin, and that she believed they would pass away if she were sexually satisfied.

Further examination showed that the patient was physically in good health. The psychological tests, including Rorschach and TAT, indicated that the patient was an immature, self-centered, impulsive individual. Sexual identification seemed particularly disturbed. There were indications of marked hostility toward men and ambivalence about her femininity. The patient was diagnosed as anxiety hysteria and treated with sedatives and psychotherapy with little immediate relief of her symptoms.

METHODS

Recording procedures. EMG's were recorded by means of an eight-channel EMG apparatus; interview content was obtained by microphone and voice recorder. Patient and therapist were in the inner shielded room of a two-room experimental suite. All recording apparatus, except for the necessary EMG leads and microphone, was in the outer room. The patient lay on a hospital bed in a relaxed position, with instructions to look straight ahead and not to move. The EMG operator observed the patient through a one-way vision window and marked the EMG tracing whenever the patient did move.

Electronic integrators. In addition to the primary EMG tracing, a second tracing, consisting of the integrated output of four EMG channels, was made on a separate recorder. The integrator tracing consisted of uniform spikes with a periodicity of 0.5 sec. The height of each spike was proportional to the EMG amplitude for each 0.5-sec. period, and could be converted directly into microvolts by reference to a calibration signal. Further details of these integration techniques have been described elsewhere (2).

Synchronization. Second-to-second synchronization of the EMG with the voice record was provided. A brief 1,000-cycle tone, audible only on the voice record, was introduced every 30 sec., while simultaneously causing a marker pen deflection on both tracings.

Electrodes. Recording was from skin electrodes with placements of the standard type used in this laboratory (2). The placements were to the right and left forearm extensors and to the right and left legs.

Measurements. Because of the lengthy records, sampling techniques were essential in measurement. For each half-minute of recording, one 2-sec. sample was measured. Criteria for measured samples were as follows: (*a*) patient talking throughout the 2-sec. period; (*b*) doctor silent for at least one sec. previously; (*c*) no gross movements taking place; (*d*) sample taken as close as possible to the half-minute signal, as a method of providing systematic randomization.

Treatment of Data

As in the previous investigation (6), the two themes singled out for study were those designated "sex" and "hostility." Interview content was classified from the typescripts. Two raters, each with full knowledge of the case, made ratings for that part of the typescript which coincided with the part of the tracing sampled (every 30 sec. as described in the preceding paper [6]). Ratings were done independently in each case, and only those items on which both raters agreed were used in the statistical analysis. Content was considered positive for "sex" when the patient was talking about actual sexual experiences or a person with whom she had had an erotic relationship. Ratings of hostility included, among other things, all references to her sister. There were 37 hostile and 58 nonhostile statements; and 45 sexual opposed to 56 nonsexual items.[3]

[3] It should be noted that the category "hostility," for example, includes items which were ambiguous with respect to the second comparison: that is, the two raters could not agree as to whether such a statement should be classed as "sexual" or not, although they did agree that it reflected hostility. The same applies to each of the other categories; and hence differences in the sample size may be expected.

TABLE 1
MUSCLE TENSION DATA IN RELATION TO CONTENT
(mean tension in microvolts)

Content	Muscle Area			
	Left arm	Right arm	Left leg	Right leg
"Hostility"	21.77	35.70	10.64	3.31
"Nonhostility"	17.02	40.80	10.95	3.54
Difference	4.75	−5.10	−0.31	−0.23
Critical ratio	1.99	1.30	0.39	0.38
Probability	.05	.18	.70	.70
"Sex"	18.45	37.90	12.08	3.56
"Nonsex"	18.45	43.70	9.73	3.68
Difference	0	−5.80	2.35	−0.12
Critical ratio	—	1.61	3.26	0.20
Probability	—	.11	.01	.84

Hostile statements were then compared with nonhostile and sexual with nonsexual, with respect to muscle tension in each of the four channels. Mean microvolt values, together with tensional differences and the corresponding probability levels are given in Table 1. When the "hostility" dichotomy is examined, the only significant difference is found in left arm tension; here hostile statements are associated with higher muscle tension. Again, for the separation with respect to statements about sex, the difference appears only in the left leg, where sexual items and higher tension go together.

DISCUSSION

This is the fourth successive case in which specific forearm tension has been related to the topic of "hostility." Without exception, all cases showed greater forearm tension associated with hostility. Three of the four cases showed the correlation of sex content with specific leg tension, and—interestingly enough—all were women. Whether specific leg tension associated with sex content is found only in women is a question for further investigation.

In the present case, it was the left forearm and left leg which showed the difference. The unilaterality in this case agrees with previous findings. Why this right-handed case should show the phenomenon in the left limbs, though, is not clear. It may be noted that Case 2 in the Shagass and Malmo study, who also was right-handed, likewise showed predominant tension in the left forearm associated with "hostility" (6, p. 306).

Theoretical Considerations

These specific muscle tensions seem to be physiological manifestations of conflict of a particular kind. Arm tension, for example, appears to be associated with conflict over hostility. Why conflict? Wouldn't it be simpler to invoke the old "law of ideomotor action"? Hostility and the impulse to hit produce a miniature striking movement, which is recorded by the sensitive EMG apparatus.

Shagass and Malmo (6) found evidence which clearly opposed this simpler explanation. This evidence was obtained from Miss A (their Case 1), the patient who was most intensively studied. When she talked about her family in interview, she was often overtly hostile in her remarks. Very few of her memories about her family could be classified as pleasant, too few in fact for statistical treatment. But there were frequent statements about members of her family which could be called neutral and some which could be interpreted as covert hostility. Arm potentials for such nonovertly hostile statements were actually higher than for overtly expressed hostility. This finding suggested that it was ambivalence or conflict over hostility (rather than conscious hostility) which was the main determinant of the higher forearm tension (6, p. 311).

Hypothesis of Central Neural Interaction

This hypothesis was actually derived from some current work on motor conflict (4). A concrete example will serve to make clear what we mean by motor conflict. A painful stimulus was applied to the arm with instructions to the subject not to move. Conflict was between the impulse to withdraw the arm to avoid pain and the desire to please the examiner, who had instructed him not to withdraw the arm. EMG data indicated that certain subjects, undergoing motor conflict, showed high arm tension when the conflict was over arm movement, and high neck tension when the conflict was over head movement (painful stimulation of the forehead). The high muscle potentials were in no way related to actual movements.

How does motor conflict produce this increase in tension? Neurophysiologically, increased tension must mean more neural impulses coming down the final common path. Very briefly, the hypothesis is this: The increase in the number of neural impulses in motor pathways is due to central neural interaction. Motor mechanisms for approach compete with similarly organized mechanisms for avoidance. This competition produces a two-way interference which evokes impulses which

are not part of any organized motor pattern. These extra or disorganized impulses going to the arm muscle would raise its tension. The same hypothesis may be applied to conflict over hostility where the opposing instigators may be—say—aggression versus affiliation or deference. A similar principle is applicable also to sexual conflicts. In a provocative article on the interaction of simultaneous responses, Meyer (5) introduced the construct of *efferent neural interaction*. Although he was not mainly concerned with conflict, much of his discussion is relevant to the present problem. There is obviously a close relationship between Meyer's *efferent neural interaction* and the notion of *central neural interaction*. The latter may be preferable because, for conflict at least, the interaction involves much more than the mechanisms which can be clearly and unambiguously designated "motor."

REFERENCES

1. DAVIS, F. H., & MALMO, R. B. Electromyographic recording during interview. *Amer. J. Psychiat.*, 1951, **107**, 908–916.
2. DAVIS, J. F. *Manual of surface electromyography.* Montreal: Laboratory for Psychological Studies, Allan Memorial Institute of Psychiatry, 1952 (mimeographed).
3. MALMO, R. B., SHAGASS, C., & DAVIS, F. H. Symptom specificity and bodily reactions during psychiatric interview. *Psychosom. Med.*, 1950, **12**, 362–376.
4. MALMO, R. B., WALLERSTEIN, H., & SHAGASS, C. Headache proneness and mechanisms of motor conflict in psychiatric patients. *J. Pers.*, 1953, **22**, 163–187.
5. MEYER, D. R. On the interaction of simultaneous responses. *Psychol. Bull.*, 1953, **50**, 204–220.
6. SHAGASS, C., & MALMO, R. B. Psychodynamic themes and localized muscular tension during psychotherapy. *Psychosom. Med.*, 1954, **16**, 295–313.

Received May 16, 1955.

A CASE STUDY OF A MULTIPLE PERSONALITY

V. K. ALEXANDER

Union Christian College, Alwaye, South India[1]

SOOSAN, a young Indian girl, aged 15 years, was brought to me by her father in May, 1952, with the complaint that Soosan would go away from home without any warning and that she was possessed with two evil spirits. One of the evil spirits was that of a young man who had committed suicide in her neighborhood, whose name was Kotchu. His suicide happened when Soosan was three years old. The other spirit was that of a grand aunt of hers, who died years before the birth of Soosan. Soosan had complete amnesia for her running away from home and for her secondary personalities. The change of personality would occur once or twice a week. The secondary personalities were so well-developed and complete in themselves, that when they came the facial expression and tone of Soosan's voice would change so much that ordinary people would regard them as different personalities. For instance Soosan was eating an unusual quantity of food at these times. Her father explained this that she needed food for three people.

Soosan had gone up to the fifth grade in school, but had discontinued her studies because of her disorder. From the onset of her symptoms, Soosan had a bitter feeling toward her younger sister, whom she tormented so much that her parents were forced to send that young child away to live with grandparents. Soosan also lacked normal affection toward her other siblings. She was the fourth child, having one elder sister and two older brothers. Her elder sister is married and has a child, of whom Soosan is very fond.

When Soosan was four months old, her mother had an attack of typhoid and hence Soosan was sent to live with her grandparents for some time. In India weaning usually takes place at the time the baby is between one and one-half and two and one-half years old. During this period Soosan had an attack of rickets. Her mother gave birth to her younger sister, when Soosan was two and one-half years old. Soosan was usually calm and quiet. She took regular part in household duties except when her personalities changed. Then she would become mischievous and use abusive language. Her father comes from a middle class family.

The method of therapy to help Soosan was twofold. First, Goddard's study (4) gave me a working hypothesis. Later I became familiar with the work of Prince (5). I wanted to know which part of Soosan's personality was dissociated. The best method, I thought, was to study the nature and behavior of the secondary personalities which were well-developed. Second, since she had a strong amnesia for these secondary personalities, I felt that it would be wise to help her build the dissociated selves. In order to do this, the easiest method, I thought, was to revive the forgotten experiences of Soosan, as Goddard had done in his case.

I used hypnosis to bring up the dissociated personalities. Soosan went readiiy into deep hypnosis. This made it easy for me to study the nature and behavior of the secondary personalities which appeared at my command. For therapeutic purposes, I also used post-hypnotic suggestion to induce Soosan to recall the events that occurred in hypnosis and to recall particularly those events occurring while she was exhibiting the secondary personalities.

As already mentioned, there were two well-developed secondary personalities, Kotchu and Anna. Of these, Kotchu was the dominant one and Anna was submissive to him. During the first session, Kotchu was asked to come. He came, as evidenced by change in Soosan's posture and gestures, but he appeared to be dumb. He refused to speak even though I tried to persuade him. At the end of the session, I gave him a paper and pencil and asked him to write down the answers. He did so very reluctantly. He wrote that his name was Kotchu and that he was dumb because he could not move his tongue, having committed siucide by hanging himself on a tree. He wrote that he hated me and did not want to come again.

[1] The writer is at present a fellow in Psychology at Princeton University, Princeton, New Jersey. His regular position is that of a lecturer in Psychology at Union Christian College, Alwaye, South India. This study was made while he was in India.

ANNA, THE TALKATIVE WOMAN

Toward the end of the first session, I also attempted to contact the other personality—Anna. Fortunately she was very talkative and talked profusely about her past life, likes and dislikes. Some of the highlights of her discourse on herself and others may be quoted here.

I am Anna. I am the wife of Poicochu. I had not seen Soosan before my death. After my death, when once I came to see my house, I saw her there. She was cleansing the water-vessel, and she did it well. Then I loved her and entered her body. [Why did you love her?][3] She is very fair when she moves about in her new dress after a bath. I do not like to have a bath. When I was alive, I used to take a bath very rarely. Since my death, I have not taken a single one. I hate it intensely. Once when I went to Soosan's home I looked at the lobsters. Then Aley, her mother, spanked me asking why I had come there. Then I went back to the cemetery through a hole in the kitchen. I live at the cemetery. . . .

I was married to Poicochu at Vakathanam. [At what age?] I do not remember. I do not love my husband. He used to beat me with household utensils. We lived together until my death. He also died later. Now when we meet, he goes away from my sight. [Why?] Oh, just to avoid a quarrel. . . .

[Tell me something about Kotchu.] Kotchu had asked me not to tell you anything about him. He also wants me not to come when you call again. [Then why did you come?] Because you wanted me to come. [What is your attitude towards me?] I do not know who you are. I heard about you just a week ago when Soosan's father said that I would be driven out by you. [Then what did you think?] I said to myself that I would not obey you. I do not want to go. [I assured her that she would not be driven out. There was a long silence, and then I asked, "What are you thinking now?"] I think that you are going to kill me. I am afraid of you. . . .

Kotchu and his brother were not on good terms. They used to quarrel. Once Kotchu told me the following. "Brother Simon kicked me once after locking me in a room. He slapped me and kicked me. The youngest sister saw all these things through the window." Kotchu can speak. He speaks to me. He says that you are a bastard. Soosan loves you, but he does not. Kotchu is afraid that he will be driven out of her body. [I assured that Kotchu will not be also driven out.] [Do you want Kotchu to be driven out?] Yes. [Why?] He is always angry. When he comes Soosan is also angry. Otherwise Soosan is a nice girl. He has warned me not to tell you these things. He is a nasty fellow. He is bad. His mind is full of lewd things. He was in love with his elder sister-in-law. They did dirty things. . . .

Soosan eats a good deal. It is Kotchu who is eating away the whole thing. She would put it on the ground [smiles] and eat it like a dog [smiles again]. There is another thing. One day when Soosan went to the attic chamber, Kotchu asked her to answer the calls of nature there. She did it (loud laughter). . . .

Soosan powders her face, but I don't. Why should one do it at the age of forty-five. . . .

Kotchu had told me that Soosan is a very bad girl. "How many days did I wait for her, but she would not come." [For what?] I do not know [A mischievous smile]. I love Kotchu more than anybody else. [Do you love Soosan?] Yes, I love Soosan too. No, I love Soosan more than Kotchu. [But you said that you love Kotchu more.] I tell you that I do not love him [Again the mischievous smile and a change of subject]. . . .

My husband and I were not on good terms. Once I went to fetch water. I put my nasal fluid into the water. My husband saw it and spanked me [laughs aloud]. Do not tell anybody. Once, while cooking food, I added chicken-casting to it. He saw it and spanked me again. He is such a fellow. Then I curse him. Now I can not help laughing when I recall these doings. . . .

When her mother spanks Soosan, it is I who experience the pain when I am in her body. Do you know why? Then I go away from her. . . .

Such exchanges went on for days, and they gradually unfolded Anna's nature and character. Anna is very dirty in her habits and appearance; dislikes taking a bath and cleaning her teeth; hates changing her dirty clothes. She was untidy in cooking and was not on good terms with her husband. In short, she is the incarnation of all that is dirty and ugly. This may represent the repressed anal stage of Soosan's development.

KOTCHU, THE INCARNATION OF EVIL

Anna was a great help in studying about Kotchu. She said that Kotchu had only one desire, i.e., to "spoil girls." He entered the body of Soosan to do "dirty things" with her. It seems that some times she likes Kotchu and at other times she dislikes him. She likes him because of his amorous nature and dislikes him because he eats everything.

In the early days of therapy, Soosan had a very interesting dream. She dreamed that Kotchu, a dark man with a white dhoti (cloth), came to her at night. She asked him, "Why did you come at this hour of the night?" Kotchu said, "I have some thing to tell you. You should not go to the Asram for treatment in future. I will never go out of your body. I will take you out of your home and 'spoil you.' Tell that bastard Alex (the therapist) to mind his own business. I will cause him to be hanged just as I was. He does not know me or the members of my family."

The dream appeared to have been a clear indication of a strong tension and resistance in Soosan.

[3] I am using brackets to indicate what I, as the therapist, said in these exchanges.

Kotchu pretended to be dumb and there was no way of studying him directly. But Soosan's parents told me that Kotchu had spoken. Anna also told me the same. Hence we thought that dumbness was due to a strong resistance and that Kotchu never wanted to reveal himself.

With Anna a good positive transference was developing. She said that she liked me and became very friendly. She gave me a useful hint. She said that Kotchu had two stages, one stage before committing suicide and the other after it. "When Kotchu is in the first stage," she said, "he can speak." I guessed from this revelation that the ice was breaking, and that these two stages were nothing but a face-saving mechanism of Kotchu himself.

We asked Kotchu of the first stage to come. He came and began to speak. At first he was a bit reserved and said that he did not like me. But after a few days there was a gradual change in his behavior. He began to talk about himself, his likes and dislikes, without any reserve. Later he became very playful even to the point of flirting. Some time he would come even without calling and would refuse to go. When he came he would take my hand and place it upon Soosan's face and kiss it or bite it playfully. Once he told me that he would like to enter my body. Once when he was in such a mood, he asked my permission to frighten me. When I agreed, Soosan rolled her eyes, ground her teeth, and made a squeaking sound in the playful Kotchu role. This assured me of a good positive transference.

We spent days permitting Kotchu to express himself freely and frankly. He had only one subject to talk about, namely, sex. As Anna pointed out, he was thoroughly amorous in nature. He wanted to seduce all the women in that locality. He took possession of Soosan's body just to enjoy her sexually. Following his death, he has continuously pursued beautiful women. He likes them. He approaches them behind and takes possession of their bodies.

Kotchu is as black as coal in color. He did not like me, the therapist, at first, because I am fairer than he. He has a body and he is a man. Once I made a slip and addressed him as a woman. He laughed aloud and said, "Is it not funny that you address me as a woman!" Once I purposely addressed him as a woman. He laughed again and retorted in an irritated tone: "You are a dog who can not distinguish between a man and a woman." At another time he covered his face and said that he was shy. I reminded him that only girls are shy. Then again he became irritated and said: "Don't you have two eyes? Look at me. My hair is cropped. I have a shirt and a shorts. My face is shaved and has three whiskers. Do women have all these things?"

Kotchu was completely antitheistic. He did not like any one to pray or go to church.[3] He hated the religious aspirations of Soosan and her friends. To him, prayer was a sign of immorality. In short, Kotchu was the incarnation of all that is evil, a kind of incarnation of the Biblical Satan. It was interesting that Soosan was terribly afraid of snakes while Kotchu was fond of them. Kotchu tells Soosan in her dreams not to be afraid of snakes. In her relaxed state, Soosan said the following:

> I am terribly afraid of snakes. I am not afraid of a biting dog, mad man or even a tiger. But I can never think of a snake; it is too frightening. . . .

The following are her free associations on snakes:

> I am afraid of it. It crawls forward. It is long. It has a small head. Once I saw a rat snake. I was alone. I cried. I saw one in front of Simon's house. Once I was standing in the place where Kotchu committed suicide. There is a pond. When a stone fell in it, I became terribly frightened, recalling that it was the place where Kotchu committed suicide. . . .
>
> It bites. It gets into a hole. Some one would kill it. There are three snakes in our well. Three together. . . .
>
> Once when I was standing on the road side, drying grain, I saw something crawling like a snake in the air. It was not very high or low. It was at the height of your waist. It was white, long, not very thick. That reminds me of snakes. . . .
>
> What will I do if it comes and plucks some of my hair. Suppose it bites me, and it drags me away. Suppose it strangles my throat and takes me away. Suppose it paralyzes my hand and after killing me, seduces me. . . .

It seemed clear that in the case of Soosan, the snake was a symbolic expression of the male sex-organ.

---

[3] Soosan and the writer are both members of an ancient and indigenous Christian church in India called the Orthodox Syrian Church. Its origin in India is traced to 52 A.D. According to the tradition, it was St. Thomas, the Apostle of Jesus Christ, who founded the church. The current Christian membership of the church comprises about one third of the total population of the Indian State of Travancore.

The following extracts taken at random, gives some further insights into Kotchu's nature and personality.

Kotchu has only one thought, that is enticing women. I will tell you one incident. Once I saw a beautiful girl, Mary. I went after her and touched her shoulders. She fell down and then I entered her body. When her relatives were trying to fix me on a tree, I escaped. I am a clever fellow. . . .

I loved the wife of my elder brother. My sister-in-law would give me anything I liked. She would give me food and something else. She gives me the 'something else' when my brother was not at home. [What is it?] It is a thing. Do you want to know? How can I tell that to you. [In a very low tone] Will you ridicule me? It is a thing which is usually asked for in the night. Don't you know what it is? Try to guess it. Please don't think ill of me or the members of my family. It is six years since she died. . . .

This Soosan has a friend named K. I do not like her. She does not behave properly. She goes to a nearby hill to pray. Is that not a sign of immorality? Once I asked her if she expects to see a young man there. Once in a dream I asked her father: "Why don't you give her in marriage?" I also asked: "Why don't you give her to me?". . . .

She always prays that what happens to her friends may not happen to her. She prays to have it only after her 25th year. She asks whether K. has it. We men do not have it. [What do you mean?] Oh, it is menses. I do do not know more about it than you do. . . .

Still another interesting aspect of Kotchu, which he revealed after some days, was his hatred of a paternal uncle of Soosan, named John. Kotchu and John are of the same age. Kotchu hates John because according to him, John has an abnormal love for Soosan and Soosan returns it. John is a bachelor aged 34 and is in military service.

We traced this matter with Soosan in her relaxed state, and it was found that she did have an abnormal love for her uncle John which was repressed. During the course of the treatment, a very interesting incident happened. Once when Soosan was under hypnosis, we asked Kotchu to come. At that moment, she opened her eyes and said: "a dark man is coming. I am afraid. He stopped at the gate of our house. He is a soldier and his cap on. [Raising her voice and changing to a cheerful tone] Hello, it is my Uncle John. How did he come here now. He went to his place of duty only last week."

After some time, Soosan reported that this man had disappeared. We asked Kotchu to come again. He came and said boastfully: "You know, it was I who came as John. I disguised myself. I am a clever fellow."

Weeks of close study convinced me that Kotchu was a personification of Soosan's sexual impulses which were repressed. The only cause to which we can attribute this repression is an oedipus situation, the father image in this case would be her uncle. In this culture, however, the paternal uncle is usually the father image and children address him as "little father."

In the beginning Soosan came for treatment every day for an hour, later once every second day, and finally, once every third day. Gradually a marked change took place in Soosan's disorder, and within a relatively short time, she showed signs of recovery. The first sign of recovery was her return to a normal diet, and toward the end she became quite well integrated. Even after three years, the secondary personalities have never reappeared and no relapse has been reported. She re-entered school and has progressed well in her studies.

Discussion

I want to draw attention to the question of dissociation. As I pointed out earlier, I worked with the definite hypothesis of dissociation, i.e., the personality was divided into different bits; the secondary personalities growing up in the unconscious mind. But after reviewing the case, I strongly feel that cases of multiple personalities are not cases of dissociation, but rather cases of repression and identification (1, p. 221; 2, pp. 436–472; 3, p. 119).

Here in the case of Soosan, the repressed oedipus situation appears to have created a tremendous tension in her life and in her ego. In order to overcome the tension, she identified herself with the secondary personalities about whom she had heard from other sources. This saved her from social as well as the super-ego's blame because the responsibility was displaced from her ego to these identified personifications of her repressed patterns. Hence the hypothesis is that Soosan's behavior is a defense mechanism of the ego to cope with her situation. This is partly determined by the cultural background also. In a culture which believes in evil spirits, it is easy for the ego to use this defense mechanism. That may explain why there were so many cases of multiple personality in New Testament times, as one can naturally assume from the description of those spirit-possessed patients whom Jesus is reported to have healed, and a few cases in the

twentieth century. As the belief in evil spirits become less and less, this type of defense mechanism becomes unacceptable both to the superego and to the society. It is little wonder that cases of multiple personality occur even now where people believe in evil spirits. It is interesting to note that in the case of multiple personality reported by Thigpen and Cleckley (6) practically no great difference was found between the various secondary personalities in EEG and Rorschach. For it is the same ego that acts in disguise in the various personalities.

REFERENCES

1. FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1948.
2. FREUD, S. A neurosis of demoniacal possession in the seventeenth century. *Collected papers.* Vol. IV, Ch. 24. London: Hogarth Press, 1946.
3. FREUD, S. *An outline of psychoanalysis.* New York: Norton, 1949.
4. GODDARD, H. H. *Two souls in a body?* New York: Dodd, 1927.
5. PRINCE, M. *The dissociation of a personality.* New York: Longmans, Green, 1906.
6. THIGPEN, C. H., & CLECKLEY, H. A case of multiple personality. *J. abnorm. soc. Psychol.*, 1954, **49**, 135–151.

Received December 30, 1954.

CRITIQUE AND NOTES

THE USE OF PERSONAL PRONOUNS AS CATEGORIES FOR STUDYING SMALL GROUP INTERACTION

DOROTHY CLIFTON CONRAD[1] AND RICHARD CONRAD

Philadelphia — *Drexel Institute of Technology*

ONE of the strongest currents in contemporary psychological thinking concerns the functioning of language in individual human behavior (3, 4). In the present paper, on the other hand, a thoroughly objective and psychologically meaningful linguistic method of exposing significant dimensions of small group interaction[2] will be described and illustrated.

This method has its origin in the assumption that the individual's pronoun choices[3] express something about his orientation toward persons, ideas, feelings, and objects. An individual's total verbal behavior within a given time unit can be categorized in terms of a pattern of pronominal preference. Under conditions of group interaction patterns of pronominal preference may be expected to express characteristics of the group relationships. The present method includes both nouns and pronouns in a common categorization.

THE METHOD

The following pronoun categories are used for the classification of pronouns and nouns, to reveal the pattern of pronominal preference. The samples provide ostensive definitions of each category. Sample nouns are enclosed in brackets.

Category	Samples
I	I; my; me; mine; myself
You	you; your; yours; yourself; [Mr. Brown]
One	one; one's; oneself; (also *you* and *your* when used in this sense, as, "You have to produce your evidence.")
He	he; his; him; himself; [Mr. Brown; Mr. Brown's]
She	she; her; hers; herself; [Mrs. Green; Mrs. Green's]
It	it; its; itself; [the book; the book's]
We	we; our; ours; us; ourselves; [this group]
You	you; your; yours; yourselves; [Mr. Brown and Mrs. Green]
They	they; their; theirs; them; themselves; [writers; books]

To assess reliability, the hypothesis was tested that the difference in assignment of pronouns to categories by the two judges is no greater than the difference to be expected from the assignment of two samples of pronouns to categories by one judge. Application of a chi-square technique[4] to a data sample yielded no evidence to require the rejection of this hypothesis.

APPLICATION OF THE METHOD

Data from two groups with distinctly different task orientations and personnel will be used to illustrate the application of the method.

Group A

A series of nine weekly meetings of a therapy group at the Veterans Administration San Francisco Regional Office Mental Hygiene Clinic constitute the data for Group A. Composite patterns of pronominal preference for this therapy group are shown in Table 1. Highest, lowest, and median percentages for each pronoun category taken separately are reported. General consistency of position for all categories is readily apparent.

It may be seen from Table 1 that most remarks made by members of the group concerned themselves as individuals. These remarks were supplemented and supported mainly by the use of the impersonal "It" category and of references to people and things as expressed in the third person plural. Remarks about individual persons typically concerned males. Females were rarely mentioned. It sometimes happened that comments were addressed directly to individual group members ("you," singular), but there was almost never any attempt by a member to separate himself from all the others by using the "you," plural category. References to the group as a whole ("we") were also rare.

Table 2 shows specific patterns of pronominal preference for each member present at a specimen

[1] Clinic data were collected while the senior author was clinical psychologist at the Veterans Administration Regional Office Mental Hygiene Clinic in San Francisco.

[2] The interaction process analysis of Bales (1) does not make use of structural linguistic categories but requires judgments of certain qualitative aspects of acts.

[3] Mowrer (2) discusses a group of unpublished studies from the University of Illinois and concludes, "... the two types of measures which are most likely to prove immediately practicable in therapy research are differential frequency of pronouns and differential frequency of tenses."

[4] In the formula $\chi^2 = \frac{(a_1 - a_2)^2}{N}$. a_1 = the number of items placed in a category by Judge No. 1; a_2 = the number of items placed in the same category by Judge No. 2; and N = the sum of a_1 and a_2. This procedure was suggested by W. L. Williams, Jr., and E. E. Cureton, both of the University of Tennessee.

TABLE 1

PATTERNS OF PRONOMINAL PREFERENCE IN NINE MEETINGS OF THERAPY GROUP

Category	Highest	Lowest	Median
I	38%	22%	32%
You	13	4	10
One	10	3	4
He	18	4	9
She	9	0	3
It	26	10	20
We	11	3	3
You	1	0	0
They	27	10	16

Note.—Number of participants in Group A: highest, 7: lowest, 4; median, 6. Total frequency of items: highest, 922; lowest, 370; median, 604.

meeting. An obvious characteristic is the extreme variability of behavior among the individual members. In terms of frequency of recorded items, for example, the variation was from seven items for Mc to 455 for Locke. Distributions for the separate categories also show decided interperson variation. Such a high degree of variability suggests that this is not an equalitarian group.

Locke clearly outstrips any other member in production of scoreable items. Is there any evidence that he influences the others? The relative frequency of remarks in the various categories by all others combined is strikingly similar to the distribution of Locke's remarks (see Table 2). No such similarity appears when the same comparison is made with respect to other individuals in the group. These data point to Locke as the member-leader of the session.

Group B

Data for Group B consist of records from two staff meetings at the same Veterans Administration clinic, at which the theory of group therapy was discussed. The designated and active leader of both meetings was the chief psychiatrist, and both meetings were attended by about the same staff members.

Results for the first meeting recorded (Table 3) showed very close similarity between the distributions of the leader's remarks and of remarks by the rest of the group. The leader himself was relatively consistent in his behavior in the two meetings, but the group members produced quite different patterns. There was a decided shift on the part of the staff to an impersonal form of communication. Relative total contributions of leader and group were almost identical for the two meetings.

The leader and individual group members unanimously reported strong feelings of satisfaction with the process and outcome of the first meeting,

TABLE 2

PATTERNS OF PRONOMINAL PREFERENCE FOR EACH MEMBER PRESENT AT SPECIMEN MEETING OF THERAPY GROUP

Category	Therapy Group Members												All Members		All Members Minus Locke	
	Jo		San		Wil		Mc		Kil		Locke					
	f	%	*f*	%	*f*	%	*f*	%	*f*	%	*f*	%	*f*	%	*f*	%
I	11	39	24	57	53	26	4	57	35	56	134	30	261	33	127	37
You	3	11	3	7	15	8	0	0	6	10	53	12	80	10	27	8
One	1	4	1	2	8	4	0	0	2	3	14	3	26	3	12	4
He	1	4	4	10	30	15	0	0	7	10	55	12	97	12	42	12
She	0	0	0	0	4	2	0	0	1	1	38	9	43	6	5	2
It	6	21	5	11	42	21	0	0	6	10	52	12	111	14	59	17
We	3	10	0	0	3	1	1	14	0	0	22	5	29	4	7	2
You	0	0	2	5	2	1	0	0	0	0	3	0	7	1	4	1
They	3	11	3	7	43	22	2	29	6	10	74	17	131	17	57	17
All categories	28	4	42	5	200	25	7	1	63	8	445	57	785	100	340	43

TABLE 3

PATTERNS OF PRONOMINAL PREFERENCE FOR LEADER AND OTHERS IN TWO PROFESSIONAL THERAPY STAFF DISCUSSIONS

Category	Discussion 1		Discussion 2	
	Leader (*f*-46)	Others (*N*-12) (*f*-58)	Leader (*f*-37)	Others (*N*-11) (*f*-47)
I	22%	28%	16%	2%
You	19	21	24	4
One	0	0	0	0
He	2	3	3	15
She	0	0	0	0
It	22	19	24	55
We	13	22	21	13
You	11	0	6	0
They	11	7	6	11
All categories	44	56	44	56

while there were feelings of disharmony and dissatisfaction with respect to the second. These observations seem predictable from the data of Table 3.

DISCUSSION

One of the basic questions that can be asked about any discussion group is whether it has developed a characteristic behavior structure. The recurring pattern of pronominal preference found in Group A seems a promising index of such a structure. Until a group "finds its way," one would expect much greater variation in patterns from meeting to meeting.

It may also be possible to discover optimal patterns by which a group can carry out various tasks. In our illustrations, each group had a different task and each followed a different pattern of

pronominal preference. The adequacy with which a group is carrying out its task might be revealed by pronominal preference analysis. For example, a therapy group with a very low percentage in the "I" category might be judged as not genuinely engaged in therapeutic activity, as a result, perhaps, of avoidance of therapeutic involvement or of the achievement of therapeutic goals.

Finally, the method potentially contributes to an understanding of the processes which determine the pronominal pattern characteristic of a group.

REFERENCES

1. BALES, R. F. *Interaction process analysis*. Cambridge. Mass.: Addison-Wesley, 1950.
2. MOWRER, O. H. Verbal behavior in psychotherapy. In O. H. Mowrer (Ed.), *Psychotherapy, theory and research*. New York: Ronald, 1953.
3. MOWRER, O. H. The psychologist looks at language. *Amer. Psychologist*, 1954, **9**, 660–694.
4. Psycholinguistics: a survey of theory and research problems. *Suppl., J. abnorm. soc. Psychol.*, 1954, **49**, Part 2.

Received March 11, 1955.

AN EXAMPLE OF ASSIMILATIVE PROJECTION

ELIZABETH MINTZ

New York University

WHEN an individual's perception of others is distorted by his own personal traits, the distortion frequently serves a defensive purpose, as in the psychoanalytic mechanism of projection. However, as Cameron (1) has pointed out in discussing "assimilative projection," an individual may exaggerate similarities between himself and others even when defensive needs do not come into play. Such a process reveals, in Murphy's phase, "an indeterminateness of the boundaries of the self" (3, p. 495) which need be in no way pathological.

To investigate this process, a group of children were intercepted in the lobby of a motion picture theater where they had seen the Walt Disney cartoon film *Peter Pan*. Eighty-six children of both sexes, ranging in age from 5 to 14, were asked, "How old do you think Peter Pan was?" They were then asked, "And how old are you?" Pearsonian product-moment correlations were computed between the ages of the children and the ages which they attributed to Peter Pan, with the results shown in Table 1.

Sharing imaginatively in Peter's adventures, the child presumably feels close to Peter and therefore perceives Peter as resembling himself. No defensive need seems to be involved in the perceptual distortion.

This minor experiment is also relevant to a current question in clinical psychology. Murray (2) developed different TAT cards for male and female subjects on the assumption that identification would take place more readily with a pictured figure resembling the subject. Recent experimental work (4, 5, 6) does not support Murray's expectation. If the TAT cards stimulate psychological processes which are akin to the responses elicited by a cartoon character such as Peter Pan, the present study falls into line with the recent experimental findings. As Table 1 shows, the correlation for girls between their own ages and the ages guessed for Peter Pan is actually higher than the correlation for boys, the difference between these correlations being significant at the .05 level.[1]

TABLE 1

CORRELATIONS BETWEEN AGES OF CHILDREN AND AGES GUESSED FOR PETER PAN

Groups	*N*	*r*	Mean of Actual Ages	Mean of Guessed Ages	*SD* of Actual Ages	*SD* of Guessed Ages
Boys	41	.34*	8.81	10.62	2.61	2.68
Girls	45	.68**	9.71	10.31	2.74	2.94
Total	86	.51**	9.28	10.46	2.71	2.82

* $p < .05$ by utilizing Fisher's r to z transformation.
** $p < .001$ by utilizing Fisher's r to z transformation.

REFERENCES

1. CAMERON, N. Perceptual organization and behavior pathology. In R. R. Blake & G. V. Ramsey (Eds.), *Perception, an approach to personality*. New York: Ronald, 1951. Pp. 283–306.
2. MURRAY, H. A. *Manual for the Thematic Apperception*

[1] The significance of the difference between these two correlations was appraised through utilizing the customary r to z transformation and comparing the difference between the two z's to its standard error. This method presupposes a normal distribution of both sets of variables, a presupposition which may not be met by the present data. One grossly deviant guess appeared in the case of a 6-year-old boy who guessed Peter Pan's age as 20. Without this deviant case, the r for boys would have been .53 and there would have been no significant difference between the r for boys and the r for girls.

Test. Cambridge: Harvard University Printing Office, 1943.
3. MURPHY, G. *Personality.* New York: Harper, 1947.
4. WEISSKOPF, EDITH A. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1950, **29**, 407–416.
5. WEISSKOPF, EDITH A., & DUNLEVY, G. P., JR. Bodily similarity between subject and central figure in the TAT as an influence on projection. *J. abnorm. soc. Psychol.*, 1952, **47**, 441–445.
6. WEISSKOPF-JOELSON, EDITH, & MONEY, L., JR. Facial similarity between subject and central figure in the TAT as an influence on projection. *J. abnorm. soc. Psychol.*, 1953, **48**, 341–344.

Received April 12, 1955.

RESPONSES TO A PERSONALITY QUESTIONNAIRE BY VOLUNTEERS AND NONVOLUNTEERS TO A KINSEY INTERVIEW[1]

ARON W. SIEGMAN

University of Wisconsin[2]

IN A paper in this *Journal*, Maslow and Sakoda (6) reported that college students who volunteered to be interviewed by Kinsey obtained higher scores on a self-esteem questionnaire than did students who refused to volunteer for such an interview. In light of a previous finding (5) that the self-esteem score correlates positively with unconventional sexual attitudes and behavior, they raised the question whether one can legitimately generalize from volunteers to a population which includes nonvolunteers. This question was also raised by other reviewers (2, 3, 4) of the Kinsey studies. In the process of studying perceptual correlates of various personality questionnaires, we obtained information that is relevant to this problem.

Students in Introductory Psychology and Psychology of Human Adjustment classes at the University of Wisconsin were solicited by their teachers for an interview by a Kinsey research team. In order to enhance the number of volunteers, *Ss* were told that the interviews were to be held at a place and time convenient for them. Also, a "secret ballot" system was adopted, and whether one volunteered or refused to do so was to be known only to the interviewers. After all *Ss* had indicated their response, they were informed that there was no Kinsey team on the campus, and the purpose of the study was explained. However, all *Ss* subsequently indicated that throughout the experiment they had no suspicion that there might not be a Kinsey team on the campus.

Whereas in the Maslow and Sakoda study 65 out of 151 *Ss* volunteered, in the present study only 24 out of the 95 *Ss* volunteered. The chi-square value for the difference between these two proportions is 7.24 which is significant beyond the .01 confidence level. In terms of sex, out of 31 male students 13 volunteered, but out of the 64 female students only 11 volunteered. The chi-square value for the difference between the two proportions is 5.53, which is significant beyond the .02 confidence level.

All *Ss* had previously taken a personality questionnaire which consisted of Taylor's Manifest Anxiety Scale, Wesley's Rigidity Scale, the defensiveness scale of the Kal Scale (7) and a Self-esteem Scale (7). Differences in anxiety, defensiveness, rigidity, and self-esteem scores of volunteers and nonvolunteers, of males and females, and the interaction between these two variables were analyzed by means of a 2 by 2 factorial design. Volunteers and nonvolunteers did not differ significantly in these respects.

The questionnaire contained the following single item referring to the *Ss*' attitude concerning sexual behavior: I believe women ought to have as much sexual freedom as men. Whereas 22 out of the 24 volunteers answered this item in the affirmative, only 30 out of the 71 nonvolunteers replied *yes*. The chi-square value for the difference between the two proportions is 15.88, which is significant beyond the .01 confidence level.Thus it seems that although no difference was found between volunteers and nonvolunteers in regard to basic personality characteristics such as manifest anxiety, rigidity, defensiveness, and self-esteem, the two groups nevertheless differ in their attitudes about sexual behavior. Such a finding is consistent with the point of view expressed by Allport (1) that values and attitudinal factors, which are underestimated in many contemporary theories of motivation as well as in clinical practice, are important determinants of behavior.

[1] The author wishes to thank Dr. H. A. Page of the Psychology Department at the University of Wisconsin for his generous help. The Self-esteem Scale discussed in this paper was developed by the author in the Psychology Laboratory of the VA Hospital, Bronx, N. Y., and he wishes to express his gratefulness to Dr. R. S. Morrow, Chief Clinical Psychologist, for his encouragement and help.

[2] Now at Teachers College, Columbia University.

Moreover, the facts that more males than females volunteered, and that on a Midwestern campus fewer volunteers were obtained than in a metropolitan eastern university, suggest that volunteers and nonvolunteers may differ in the very sociocultural variables which Kinsey found to be sources of variation in sexual behavior. In comparing the proportion of volunteers obtained by Maslow and Sakoda, with that obtained in this study, however, one has to bear in mind that the present study took place in the spring of 1954, by which time the Kinsey reports had gained considerable more publicity.

REFERENCES

1. ALLPORT, G. W. The trend in motivational theory. *Amer. J. Orthopsychiat.*, 1953, **23**, 107–119.
2. COCHRAN, W. G., MOSTELLER, F., & TUKEY, J. W. Statistical problems of the Kinsey report. *J. Amer. Statist. Ass.*, 1953, **48**, 673–716.
3. DOLLARD, J. The Kinsey report on women: "A strangely flawed masterpiece." *N. Y. Herald Tribune*, Sept. 13, 1953, Section 6, p. 3.
4. HYMAN, H., & SHEATSLY, P. B. The scientific method. In D. P. Geddes (Ed.), *An analysis of the Kinsey reports*. New York: New American Library, 1954. Pp. 93–118.
5. MASLOW, A. H. Self-esteem and sexuality in women. *J. soc. Psychol.*, 1942, **16**, 259–294.
6. MASLOW, A. H., & SAKODA, J. M. Volunteer-error in the Kinsey study. *J. abnorm. soc. Psychol.*, 1952, **47**, 259–262.
7. SIEGMAN, A. W. Some factors associated with visual threshold for taboo words. Paper read at Midwest Psychol. Ass., Chicago, 1955.

Received April 19, 1955.

THE USE OF THE MMPI FOR CROSS-CULTURAL PERSONALITY STUDY: A PRELIMINARY REPORT ON THE GERMAN TRANSLATION[1]

NORMAN D. SUNDBERG

University of Oregon

THE Minnesota Multiphasic Personality Inventory (MMPI) is widely used in America, but the possibility of its application to other cultures has received little attention. In fact, questionnaire-type personality tests in general have scarcely been employed in intercultural comparisons. Yet many cultural products, e.g., novels, are constantly being translated into other languages and successfully disseminated. Language is the most common medium of communication between people, and much of what interests the anthropologist and psychologist is expressed in language. Standardized verbal tests could serve as supplements to the unstandardized interviews frequently used in intercultural comparisons. Such tests also possess certain advantages over projective materials, which are more commonly employed in intercultural studies. With literate people large samples can be obtained by administering structured personality tests in groups. Structured tests are more amenable to statistical treatments, item analysis, etc., than projective techniques. The MMPI would seem to be especially interesting for intercultural studies because of the large amount of research on it in this country and because of the richness of its item pool, which was originally developed as a substitute for the psychiatric interview.

Translations of the MMPI into Spanish, French, and Japanese exist (2), but as yet there have been no publications concerning them except one report in a French journal (1). The present study is the initial one in a series of investigations using the first German translation.

The first problem which arises in the intercultural usage of verbal tests is the difficulty of translation. In order to get a comparable test, the flavor as well as the exact meaning of the original test must be preserved as much as possible. Consider the difficulty of translating such idioms as "to put one over on me" or "to cut up." In addition, there were certain difficulties because of differences in culture. One MMPI item has a statement about the "funnies." In Germany there is no exact equivalent of the "funnies" in the newspapers. Another item refers to Lincoln and Washington, two names not well known in Germany and certainly unlikely to carry the same connotations there as in America.

Even with a good translation, there are still many barriers to interpreting the MMPI results obtained in another culture in the same way that one interprets results from this culture. MMPI scales and patterns might turn out to have differ-

[1] This article is a revision of a paper entitled "The Performance of German Students on a Translation of the MMPI," read at the joint Oregon-Washington State Psychological Association meeting at Pullman, Washington, May 1954. The writer wishes to express his appreciation for a scholarship from Student Projects for Amity among Nations, University of Minnesota, in 1948, and a research grant from the University of Oregon Graduate School, 1953–54.

ent distributions in another population and different reliabilities and validities. The best solution might be a reconstruction of the scales and a restandardization of the test in the German setting. Such an ambitious project will have to remain for the future. In this study, the initial translation will be described and the results with a German student sample compared with data from an American student sample.

The original form of the German translation was developed in 1948.[2] Only the validity scales, the usual clinical scales, and the Social Introversion scale were translated, making a total of 399 items as compared with the total of 566 on the group form. With the test was presented a translation of the instructions for the use of IBM answer sheets.

THE GERMAN STUDENT SAMPLE

One hundred and twenty German students at the International Holiday Course at the University of Marburg, in the summer of 1948, volunteered or were persuaded to take the test anonymously.[3] The age range was from 18 to 35 years. Previous university attendance averaged five semesters with a large range from zero to eleven semesters. This average would roughly correspond to the American senior year or first graduate year, since German schools are so organized that the student enters a university at a later educational stage. About half of the German students were from Marburg University and the rest from other western zone universities. Forty per cent were students in the faculty of philosophy, which corresponds approximately to the American liberal arts college. Twenty per cent were in the faculty of law, 12 per cent in medicine, and 6 per cent in theology; the rest were in a number of smaller faculties other than the four traditional ones, or did not provide such information.

RESULTS

Those records which had 50 or more unanswered items were eliminated. The small number of records which showed deviant scores on other validity scales were not eliminated since the meaning of these scores in a German sample was not clear.

[2] Deciding it would be very dangerous for a nonnative to translate alone anything as colloquial as the Multiphasic, the writer gained the assistance of two University of Marburg students, Fräulein Ursula Ehrhardt, a sociology student, and Herr Wolfgang Lauber, a medical student. Both had studied English several years and had conversed extensively with Americans in the occupation forces. In addition, Dr. Klittscher, professor of English at the University of Marburg, checked the whole translation before it was used.

[3] The German students took the test in much the same manner as American students. They finished in the amount of time which would be expected, although they were not acquainted with IBM answer sheets. As far as the writer could ascertain, there was no personality test like the MMPI in Germany.

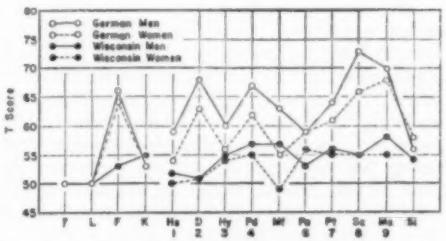

FIG. 1. PROFILES OF MMPI MEAN SCORES OF GERMAN STUDENTS USING THE GERMAN TRANSLATION AND OF WISCONSIN FRESHMEN USING THE ORIGINAL ENGLISH VERSION

The final German sample included 51 women and 60 men.

Two different methods of comparison with American college groups were used. Figure 1 shows the mean profile for the men and the women of the German sample and for the 760 women and 1,422 men of the entering freshman class at the University of Wisconsin in 1948.[4] The many large and significant differences are obvious on inspection. Critical ratios were significant at the 1 per cent level for all scales except *L* and *Hy* for women and *K* for the men. The only scale on which the German students were significantly lower was *K* for women. The German men's standard deviations are significantly larger on the *L*, *F*, and *Sc* scales but on no others.

The percentages of German and American students showing peaks on the various clinical scales were also compared. Data were derived from a tabulation of the first digit in the MMPI code, i.e., the scale with the highest *T* score, in a manner similar to that followed in the MMPI *Atlas* (3). The most notable feature was the absence of German codes without high points, the most frequent category for Wisconsin codes. All of the German profiles had one or more *T* scores over 54.

DISCUSSION

In discussing the results, it would be foolhardy to assume that the same interpretations as are applied to American profiles and codes could be ap-

[4] The Wisconsin student data were made available through the courtesy of Dr. L. E. Drake, University of Wisconsin. Tables presenting statistical comparisons between the German and American students have been deposited with the American Documentation Institute. Order Document No. 4686 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

plied to German ones. The large and significant differences might be due to a number of factors, only some of which can be even partially identified.

One of the factors to investigate is the adequacy of the translation. It is interesting that all of the significant differences are in the so-called abnormal direction. This finding suggests that the German translation might be "milder" or less extreme in implication. The German readers would then give more answers which would be scored on the various scales. At present a thorough check of the translation is being carried out.

The results may point towards real psychological differences between the German and American students. These differences might merely be differences in test-taking attitude or differences in distributions of total test scores, but they might also represent more fundamental psychological-cultural differences. If there are basic differences, one cannot yet be sure what they are. They may be related to what is ordinarily associated with the various MMPI scales, but they may not. Personal observation suggests no reason to believe that the German students suffer from more psychopathology than American students.

SUMMARY

Structured personality tests have seldom been used in intercultural comparisons, but present certain advantages over other techniques. The first study using a German translation of the MMPI has been reported. German men and women attending a summer course at the University of Marburg in 1948 showed profiles significantly higher on most scales than American students. Since there is a mixture of factors which might have effects, interpretation of the results is not clear, but the method seems to merit serious attention.

REFERENCES

1. GAYRAL, L., CARRIE, J., JARDILLIER, P., & SCIOLETTE, GENEVIEVE. Etude critique de l'inventaire multiphasique de personnalite ("Test" de Minnesota). *Ann. med.-psychol.*, 1951, **109**, 166–178.
2. HATHAWAY, S. R. Personal communication, 1954.
3. HATHAWAY, S. R., & MEEHL, P. E. *An atlas for the clinical use of the MMPI*. Minneapolis: Univer. of Minnesota Press, 1951.

Received February 8, 1955.

GROUPING ERROR AND INTERPERSONAL AFFECT: A CORRECTION[1]

EVERETT W. BOVARD, JR.

University of Toronto

PREVIOUS research by the writer (1) has suggested that verbal interaction induces positive feeling among small group members, both with respect to each other, and with respect to the group as a whole.

The former finding was based on multiple ratings from each subject (who was asked to rate every other person in the group), with N the total number of ratings rather than the number of persons making them. This inflated N led to an underestimated standard error of the mean and hence to an erroneously small estimate of the probability that the results, which were in the expected direction, could have occurred by chance. Multiple observations from each subject cannot be summated in this manner because a fundamental assumption of the probability statistics involved is that each observation is independent of every other—and this is obviously not true of ratings made by the same person, which are correlated. The error involved in violating this assumption has been termed "grouping error."

This argument does not, it should be noted, involve the finding in respect to the group as a whole, which was based on a single observation from each subject.

The data involved in this study can, however, be reanalyzed[2] to test the original hypothesis, without assuming independence of the multiple ratings made by each subject. In this method,[3] a single score is assigned to each of the N persons in

[1] The writer is much indebted to the previous editor of this Journal, Dr. J. McV. Hunt, and to an anonymous editorial consultant for pointing out the error in assuming multiple observations from the same subject are independent in a study of this kind. Grateful acknowledgement is also made to F. W. R. Taylor for the statistical computations on which the present paper is based.

[2] A further finding from another group that the average rating of every person for every other on the affect scale rose significantly following a party (1, p. 525) was also based on the erroneous assumption that multiple ratings from the same person are independent. Reanalysis of these data was not undertaken, but examination of the data suggests this difference in affect level is not statistically significant.

[3] This method of analysis was suggested by another editorial consultant to this *Journal*, to whom the writer is much obliged.

TABLE 1

MEAN RATINGS ASSIGNED GROUP MEMBERS IN GROUP-CENTERED AND LEADER-CENTERED CLASSES ON AFFECT SCALE

Leader	Mean: G/C	*N*	Mean: L/C	*N*	$SE_{diff.}$	*CR*	*p*
A	2.09	23	1.23	22	.212	4.05	<.00006
B	1.33	18	1.13	18	.207	1.00	<.32

the group, the average of the *N*-1 ratings of the person by everyone else. If not everyone else rated this person, the score assigned is the average of those who did. The *N* used to compute the mean and its standard error is thus the number of persons in the group, and not the number of ratings.

Results obtained when the original data were reanalyzed according to this procedure are shown in Table 1. The hypothesis being retested is that students in each of the two experimental (group-centered) classes, where verbal interaction was encouraged, would rate each other higher on an 11-point affect scale than would be the case, respectively, in each of the two matched control classes, where interaction was held to a minimum.

The difference between means in each of the two experimental-control pairs is in the expected direction. Since the results from each of the two pairs are independent, the combined probability that both differences could have occurred by chance can be determined by the method developed by Jones and Fiske (2). This probability is less than .001 ($\chi^2 = 22.09$ at 4 degrees of freedom), and thus the original finding that interaction induces positive feeling among group members is confirmed, albeit at a much reduced level of significance than that originally reported.

REFERENCES

1. BOVARD, E. W., JR. The experimental production of interpersonal affect. *J. abnorm. soc. Psychol.*, 1951, **46**, 521–528.
2. JONES, L. V., & FISKE, D. W. Models for testing the significance of combined results. *Psychol. Bull.*, 1953, **50**, 375–382.

Received December 9, 1955. Prior publication.

SOCIOMETRY

Founded by J. L. Moreno, in 1937

BACK VOLUMES NOW AGAIN AVAILABLE

Starting with Vol. I

Due to the continued requests for out of print back volumes of *Sociometry*, the publishers are pleased to announce that these are being made available, to replace worn out, lost or missing copies.

Clothbound in Black

$16.50 Each

Order from: BEACON HOUSE Inc.

P. O. Box 311 Beacon, N. Y.

Have You Seen . . .

Contemporary Psychology

Featured in the February issue:

The Life and Work of Sigmund Freud

If you've not yet read Volume 2 of Ernest Jones's biography, let *CP* come to your aid! Read the absorbing, five-page book review of Freud's life, as Jones portrays it, in the important years from 1901 to 1919.

- Every month *CP* presents many reviews of the important, current books in social psychology and psychotherapy. And, these are all critical, evaluative reviews prepared by specialists in their particular fields. Let *CP* help you "keep up" with the new literature in abnormal and social psychology.
- Other *CP* features include: The Film Department—here's valuable information on all phases of educational films of particular interest for teaching and research staffs in the emotional and mental illnesses; "*CP* Speaks"—the editor's page of news and comment on developments in psychology's wide world of books; and, "On the Other Hand"—a section reserved for comments by *CP*'s readers.

Contemporary Psychology belongs on your reading shelf. The subscription rate is $4.00 to APA members and participants of Student Journal Group; $8.00 to all others (add $.50 extra for foreign postage).

Send In Your Subscription Now*

While Copies of the January and February Issues Are Still Available

American Psychological Association

1333 Sixteenth Street N. W. Washington 6, D. C.

* If you wish, a letter *addressed to the attention of Dept. ASS*, indicating your interest in a subscription to *CP* and mentioning your specific subscription group, i.e., "APA member," will be sufficient. We'll then forward your subscription bill to you.